

ISSN: 0268-1242

SEMICONDUCTOR SCIENCE AND TECHNOLOGY

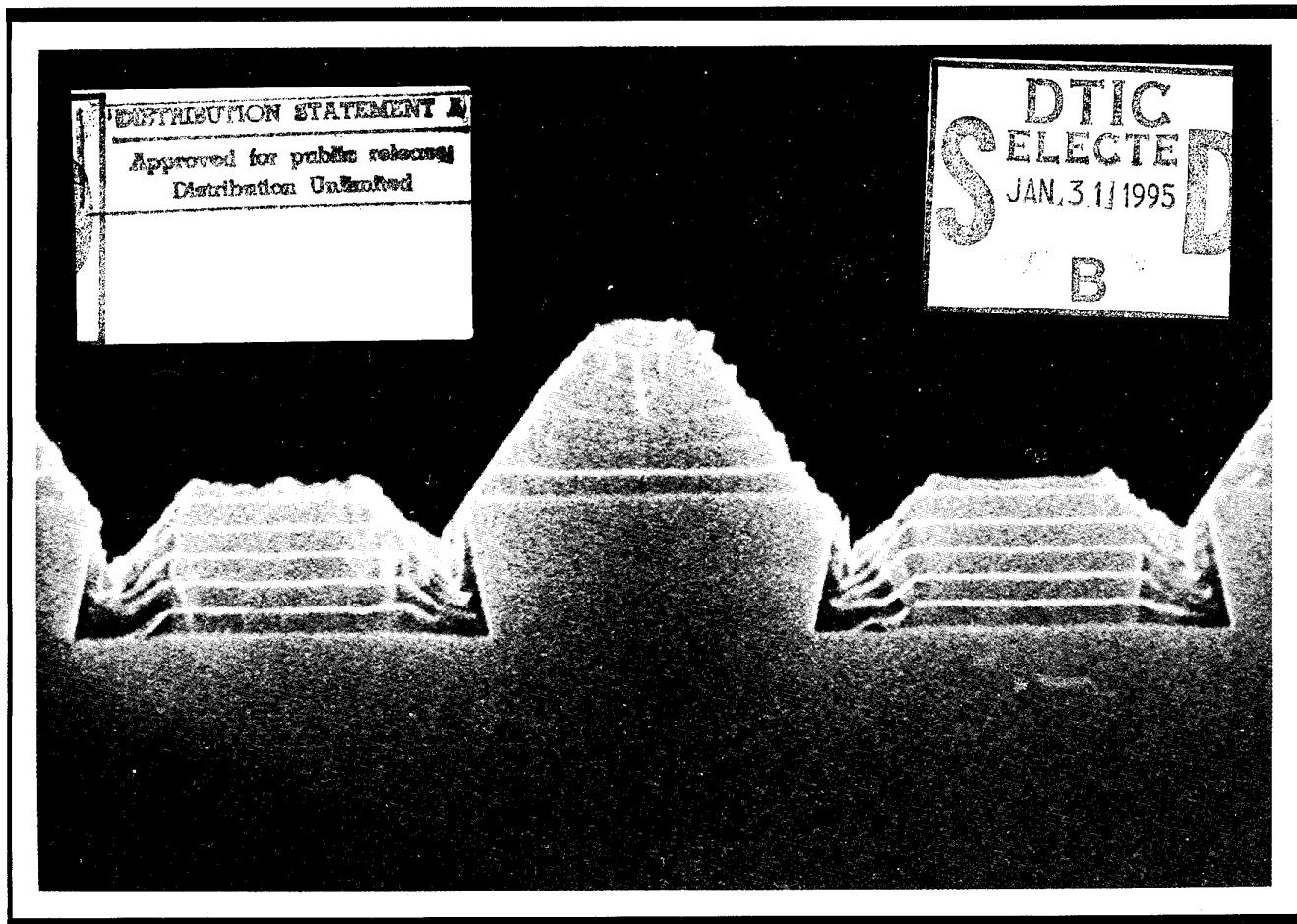
1994-01-16

8th Winterschool on New Developments in Solid State Physics
Mauterndorf, Austria, 14–18 February 1994

Volume 9

Number 11S

November 1994



DTIC QUALITY INSPECTED



Institute of Physics Publishing



A Journal Recognized by the European Physical Society

SEMICONDUCTOR SCIENCE AND TECHNOLOGY

Semiconductor Science and Technology is an international multidisciplinary journal publishing original research papers, Letters and Review Articles on all aspects of the science and technology of semiconductors. It provides a forum for physicists, materials scientists, chemists and electronics engineers involved in all the disciplines related to semiconductors, from the theoretical discussion and experimental observation of basic semiconductor physics, through the growth and characterization of samples, and the fabrication of structures to potential device applications.

Honorary Editor

E Gornik, Technical University of Vienna, Austria

Editorial Board

M Asche, Paul-Drude-Institut für Festkörperferelektronik, Berlin, Germany

G Bauer, Johannes Kepler Universität Linz, Austria

S P Beaumont, University of Glasgow, UK

M Erman, Alcatel-Alsthom Research, Marcoussis, France

C T Foxon, University of Nottingham, UK

E Göbel, Philipps-Universität Marburg, Germany

M Grynpberg, University of Warsaw, Poland

M Heiblum, Weizmann Institute of Science, Rehovot, Israel

P Lugli, Università di Roma 'Tor Vergata', Rome, Italy

J C Maan, Catholic University of Nijmegen, The Netherlands

R J Nicholas, University of Oxford, UK

R A Suris, Ioffe Physico-Technical Institute, St Petersburg, Russia

V B Timofeev, Institute of Solid State Physics, Chernogolovka, Russia

L Vina, Instituto de Ciencia de Materiales, Madrid, Spain

P Voisin, Ecole Normale Supérieure, Paris, France

G Weimann, Technical University of Munich, Germany

American Sub-Board

D G Seiler (*Chairman*), National Institute of Standards and Technology, Gaithersburg, MD, USA

S G Bishop, University of Illinois at Urbana-Champaign, USA

E Mendez, IBM Watson Research Center, Yorktown Heights, USA

U K Mishra, University of California, Santa Barbara, USA

A Ourmazd, AT&T Bell Laboratories, Holmdel, USA

T J Shaffner, Texas Instruments Inc., Dallas, USA

Japanese Sub-Board

C Hamaguchi (*Chairman*), Osaka University, Japan

N Sawaki, Nagoya University, Japan

Y Shiraki, University of Tokyo, Japan

Materials and Device Reliability Sub-Board

H E Maes (*Section Editor*), IMEC, Leuven, Belgium

A Birolini, ETH Zentrum, Zurich, Switzerland

H Hartnagel, Technische Hochschule, Darmstadt, Germany

C Hu, University of California, Berkeley, USA

J W McPherson, Texas Instruments Inc., Dallas, USA

E Takeda, Hitachi Ltd, Tokyo, Japan

E Wolfgang, Siemens AG, Munich, Germany

Publisher R Cooper

Editorial M B Taylor

Production A D Evans

Editorial and Marketing Office

Institute of Physics Publishing

Techno House, Redcliffe Way

Bristol BS1 6NX, UK

Tel: 0272 297481; Fax: 449149

Fax: 0272 294318

E-mail: Internet:

sst@ioppublishing.co.uk

X400: /s = sst/o = ioppl/

prmd = iopp/admd = 0/c = gb

Consultant Editor

V Grigor'yants

IOPP Editorial Office

Ioffe Physico-Technical Institute

26 Polytechnicheskaya

194021 St Petersburg, Russia

Advertisement Sales

Jack Pedersen/John Irish

D A Goodall Ltd

65/66 Shoreditch High Street

London E1 6JH

Tel: 071-739 7679

Fax: 071-729 7185

Published monthly as twelve issues per annual volume by Institute of Physics Publishing, Techno House, Redcliffe Way, Bristol BS1 6NX, UK

Subscription information – 1994 volume

For all countries, except the United States, Canada and Mexico, the subscription rate is £604.00 per volume. Single issue price £46.50 (except conference issues/supplements – prices available on application). Delivery is by air-speeded mail from the United Kingdom to most overseas countries, and by airfreight and registered mail to subscribers in India. Orders to:

Order Processing Department
Institute of Physics Publishing
Techno House, Redcliffe Way
Bristol BS1 6NX, UK

For the United States, Canada and Mexico, the subscription rate is US\$1238.00 per volume. Delivery is by transatlantic airfreight and onward mailing. Orders to:

American Institute of Physics
Subscriber Services
500 Sunnyside Boulevard
Woodbury, NY 11797-2999, USA

Back volumes

Orders and enquiries for the 1993 volume should be sent to the subscription addresses given above.

United States Postal Identification Statement:

Semiconductor Science and Technology (ISSN: 0268-1242) is published monthly for \$1238.00 per volume in association with the American Institute of Physics, 500 Sunnyside Blvd, Woodbury, NY 11797. Second-class postage paid at Woodbury, NY and additional mailing offices. POSTMASTER: Send address changes to *Semiconductor Science and Technology*, American Institute of Physics, 500 Sunnyside Blvd, Woodbury, NY 11797.

Pre-publication abstracts of articles in *Semiconductor Science and Technology* and other related journals are now available weekly in electronic form via CoDAS, a new direct alerting service in condensed matter and materials science run jointly by Institute of Physics Publishing and Elsevier Science Publishers. For details of a free one-month trial subscription contact Paul Bancroft at fax +44 272 294318 or email bancroft@ioppublishing.co.uk.

Copyright © 1994 by IOP Publishing Ltd and individual contributors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publishers, except as stated below. Single photocopies of single articles may be made for private study or research. Illustrations and short extracts from the text of individual contributions may be copied provided that the source is acknowledged, the permission of the authors is obtained and IOP Publishing Ltd is notified. Multiple copying is permitted in accordance with the terms of licences issued by the Copying Licensing Agency under the terms of its agreement with the Committee of Vice-Chancellors and Principals. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients in the USA, is granted by IOP Publishing Ltd to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$19.50 per copy is paid directly to CCC, 27 Congress Street, Salem, MA 01970, USA.

Typeset in the UK by Integral Typesetting, Great Yarmouth NR31 0LU and printed in the UK by William Gibbons & Sons Ltd, Wolverhampton WV13 3XT.

The text of *Semiconductor Science and Technology* is printed on acid-free paper.

Foreword

The Eighth International Winterschool on New Developments in Solid State Physics, entitled *Interaction and Scattering Phenomena in Nanostructures*, was held in Mauterndorf Castle, Salzburg, Austria on 14–18 February 1994. A total of 69 papers (including posters) were presented at the meeting. 28 invited papers are printed in this volume.

As usual, it was intended to have the most recent highlights in low dimensional physics presented at this meeting. The main topics were:

- Composite fermions and fractional quantum Hall effect
- Mesoscopic transport and chaos
- Low dimensional tunnelling
- Wires: spectroscopy and lasing action
- Bloch oscillations and ultrafast phenomena
- Coupled quantum wells and superlattices
- Si/SiGe heterostructures
- Microcavities

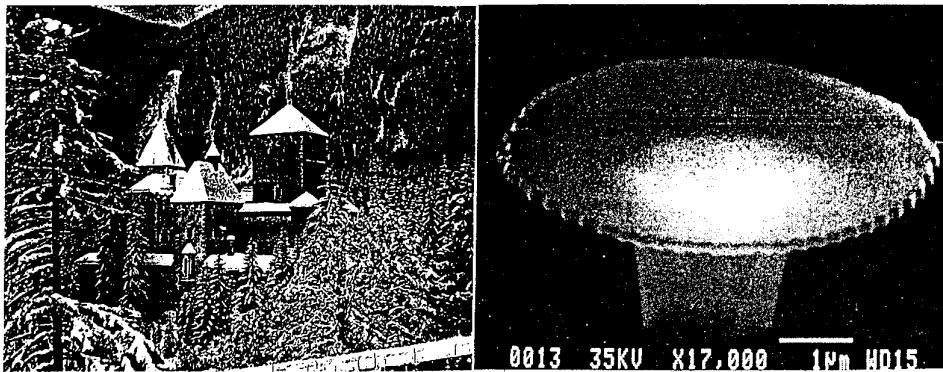
The success of this conference series relies heavily on the invited speakers, who made real efforts to give lucid presentations of their work. The event's strong international tradition was maintained by a total of about 190 scientists attending from 20 countries.

The social programme culminated in the traditional ski race, which was held on Friday afternoon. For the first time a non-Austrian team ('Raman Express' from Munich) won the race. A new dimension to the Conference was added by Wlodek Zawadzki, who read a chapter from his recently published novel '*Grand-Inquisitor*' after an evening session.

A large number of people contributed through their advice, support and collaboration to the conference: to all of them the organizers are most grateful.

G Bauer, F Kuchar and H Heinrich
Conference Co-Editors

19950125 176



**8th International Winterschool on
New Developments in Solid State Physics**

**INTERACTION AND SCATTERING PHENOMENA IN
NANOSTRUCTURES**

14-18 February, 1994
Mauterndorf, Salzburg, Austria

Organizing Committee

G. Bauer
F. Kuchar
H. Heinrich

Johannes Kepler Universität Linz, Austria
Montanuniversität Leoben, Austria
Johannes Kepler Universität Linz, Austria

Local Organization

B. Füricht
U. Hannesschläger

K. Rabeder
U. Stögmüller

I. Kuchar

The conference organizers would also like to thank the following companies for their support:

Balzers AG, Austria
Bruker Analytische Messtechnik, Germany
Coherent GmbH, Germany
Instruments S.A./Riber, Germany
Mausz Vakuumtechnik, Austria
Oxford Instruments, Germany
VTS-Joachim Schwarz, Germany

Conference acknowledges substantial support by the Bundesministerium für Wissenschaft und Forschung, Austria, Österreichische Physikalische Gesellschaft, Austria, Österreichische Forschungsgemeinschaft, Austria, Gesellschaft für Mikroelektronik, Austria, US Air Force , Office of Scientific Research, EOARD, London, UK, Office of Naval Research, USA .
This work relates to Department of Navy Grant N00014-94-1-0401 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained here in.

The fractional quantum Hall effect in a new light

H L Stormer[†], R R Du^{‡§}, W Kang[†], D C Tsui[§], L N Pfeiffer[†],
K W Baldwin[†] and K W West[†]

[†] AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

[‡] Francis Bitter National Magnet Lab, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA

[§] Princeton University, Princeton, NJ 08544, USA

Abstract. Today the fractional quantum Hall effect (FQHE) is well understood as the condensation of two-dimensional electrons in a high magnetic field into a sequence of quantum liquids. Thus far the relationship between the different liquids has been viewed as a hierarchy in which higher-order fractional states develop from the condensation of the elementary excitation of the next lower-order fractional state. Recent theory has shed a new light on these relationships and provided us with a new framework for the FQHE. The central ingredients of this new picture are bizarre new particles, often termed composite fermions. This paper reviews our recent experimental evidence for the existence of such objects.

1. Introduction

The physical picture of the fractional quantum Hall effect (FQHE) [1] that has evolved during the past decade or so can be summarized in the following way [2, 3]: at low temperatures and in high magnetic fields, electrons in two-dimensional (2D) systems, under the influence of their dominating Coulomb interaction, condense into a sequence of quantum liquids at filling factors $v = p/q$ with q always odd (disregarding spin effects). The filling factor indicates the fraction to which a Landau level is filled with electrons. The most prominent of these liquids exist at filling factors $v = 1/q$, approximated extraordinarily well by Laughlin's succinct many-body wavefunction [4]. A gap exists above this ground state leading to vanishing electrical conductivity (as well as resistivity, due to the matrix relationship between σ and ρ) at precisely such filling factors and in their vicinity, the latter caused by carrier localization. Concomitantly, the Hall effect is quantized to hq/e^2 .

Excitation across the energy gap creates quasiparticles which carry a fractional charge, e/q . The system reacts to deviations from exact filling of one of the prominent states by generating such quasiparticles which, at appropriate densities, condense into liquids of *quasiparticles*. This process can be repeated successively creating a hierarchy [5–7] of FQHE states at $v = p/q$, the so-called higher-order fractions, emanating from the primary states at $v = 1/q$. Taking into account electron-hole symmetry and translation to higher Landau levels, this hierarchy covers all odd-denominator fractions.

The description of the FQHE at general, rational filling factor $v = p/q$ has never been as satisfactory as the description of the primary states at $v = 1/q$ (and their

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification _____	
By _____	
Distribution _____	
Availability Codes	
Dist	Avail and/or Special
A-1	_____

electron-hole equivalent at $v = 1 - 1/q$). Firstly, the primary states and the derived daughter states are placed on an unequal footing, although experimentally they appear to be equivalent. Furthermore, the hierarchy orders the FQHE states like relatives on a family tree, with a different $1/q$ fraction at the root of each tree, but it does not provide a measure for their relative strength (size of the gaps). Consequently, the reason for the predominant appearance of certain FQHE sequences in experiment and the lack of other states does not find a clear interpretation within this model. And finally, whereas Laughlin's elegant wavefunction describes the primary $1/q$ states extraordinarily well, wavefunctions for the higher-order states are much more complex and less satisfactory [8–10].

While the higher-order FQHE states at odd-denominator filling factors persisted in being ill-described, the states at even-denominator fractions remained totally enigmatic (we exclude here states such as $v = 5/2$ which are true FQHE states caused by spin effects) [11, 12]. Transport features, although generally weak, around $v = 1/2$ were largely viewed as caused by the convergence [11] of a sequence of FQHE states at $v = p/(2p \pm 1)$ at the outer boundaries of the hierarchies emanating from the primary states at $v = 1/3$ and $v = 2/3$. The state of affairs for the higher-order fractions, as well as for the even-denominator states, remained less than satisfactory.

An early experimental paper by Jiang *et al* [13] clearly exposed our lack of understanding of the electronic state at $v = 1/2$. In a very high-quality 2D sample a deep minimum appeared in the magnetoresistivity, ρ_{xx} , at $v = 1/2$, the temperature (T) dependence of which was quite distinct. Different from the exponential T dependence of the FQHE states, the state at $v = 1/2$ showed a

roughly linear T dependence converging towards a non-zero ρ_{xx} value as $T \rightarrow 0$. Subsequent experiments on surface acoustic wave propagation by Willett *et al* [14] uncovered anomalies in the attenuation and velocity shifts at $v = 1/2$ that could not be explained in terms of a traditional FQHE state.

Jain [15] has shed new light on the higher-order fractions by proposing a bold generalization of Laughlin's trial wavefunction which formally invokes the wavefunction of higher-order Landau levels. This theory implies a mapping of the states at $v = p/(2p+1)$, for example, onto the electron Landau levels at integer $v = p$ (spin neglected). In this way, the states at $1/3, 2/5, 3/7 \dots$ resemble electron systems at Landau level filling factor $1, 2, 3, \dots$ respectively. However, in the case of the fractional states these 'Landau levels' are occupied not with electrons but with strange particles, termed composite fermions [15]. They consist of electrons to which an even number of flux quanta have been attached by virtue of the electron-electron interaction. Jain's wavefunctions compare very favourably with numerical few-particle calculations [16], which gives rise to intriguing speculations as to further analogies between the gaps in the FQHE and regular Landau levels, as well as to its limiting behaviour as $p \rightarrow \infty$. For large p , the sequence of Landau levels filled with electrons tends towards the Fermi liquid state at $B = 0$ while the analogous sequence of FQHE states converges towards $v = 1/2$. By way of analogy, this suggests that composite fermions at $v = 1/2$ may behave similarly to electrons at $B = 0$ and form a Fermi liquid.

This 'top-down' speculation from the FQHE states to the state at $v = 1/2$ has been complemented and strengthened by a recent 'bottom-up' proposal by Halperin *et al* [17] (HLR) which starts with a Chern-Simons gauge transformation at $v = 1/2$ and expands towards the fractions. Their theory, too, involves even-number flux attachment to electrons. The two flux quanta per electron at $v = 1/2$ are incorporated into the so-called Chern-Simons particles and the average magnetic field, therefore, apparently has vanished. The electronic state becomes a Fermi liquid of Chern-Simons fermions with a well defined Fermi wavevector k_F . As the magnetic field deviates from $v = 1/2$, the new particles experience an effective magnetic field, and their orbits are being quantized into 'Landau levels' of Chern-Simons fermions which gives rise to the sequence of FQHE states around $v = 1/2$. In this way, the gaps in the excitation spectra of the liquids at $v = p/(2p+1)$ represent the gaps between 'Landau levels' of Chern-Simons particles whose value, in analogy to the regular electron case, can be associated with the effective mass of the particle.

Intuitively, one is led to believe that the 'top-down' view starting from the $1/3$ -trial wavefunction and extrapolating towards $v = 1/2$ and the 'bottom-up' view starting from a gauge transformation at $v = 1/2$ and working towards $v = 1/3$ are merely different reflections of the same underlying states. One also would want to equate Chern-Simons particles with composite fermions and think of both of them in analogy to simple electrons in a magnetic field for which there exist well established

theoretical concepts and experimental tools. While the equivalence of the 'top-down' and 'bottom-up' approach is being addressed by theory, we may explore, experimentally, to what extent the *ad hoc* electron analogy generates meaningful statements about the states around $v = 1/2$. Towards this end one can perform several magnetoresistance experiments which are being reviewed here.

2. Thermal activation energy measurements

Despite their dominance in the FQHE, experimental determination of gap energies of higher-order states have not been widely pursued. Only recently has sample quality reached a level where gap energy measurements can be performed on several states of a given sequence. The energy gap Δ_v at a filling factor v can be deduced from the exponential temperature dependence of the magnetoresistivity $\rho_{xx} \propto \exp(-\Delta/2kT)$. Du *et al* [18] have performed such measurements on a modulation-doped GaAs/AlGaAs heterostructure of density $n = 1.12 \times 10^{11} \text{ cm}^{-2}$ and mobility $\mu = 6.8 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Figure 1 shows an overview of the FQHE in the lowest Landau level at $T = 40 \text{ mK}$. It reveals the extraordinarily high quality of the sample. The series of fractions $p/(2p \pm 1)$ converging towards $v = 1/2$ show ρ_{xx} features up to $v = 9/19$ and $9/17$. The data exhibit the common, broad minimum at $v = 1/2$. FQHE states descending from $v = 1/4$, in a series of $p/(4p \pm 1)$ are also seen in the trace. The temperature dependence of the resistivity minima at filling factor $v = 1/3, 2/5, 3/7, 4/9$ and $5/11$ is exponential and, except for the last fraction, extends over more than one decade in ρ_{xx} . From the slope of $\log \rho_{xx}$ versus $1/T$ we can deduce the energy gaps Δ_v for the fractions at filling factor v .

It is most illuminating to plot the so-determined gap energies Δ_v versus the applied magnetic field. Figure 2

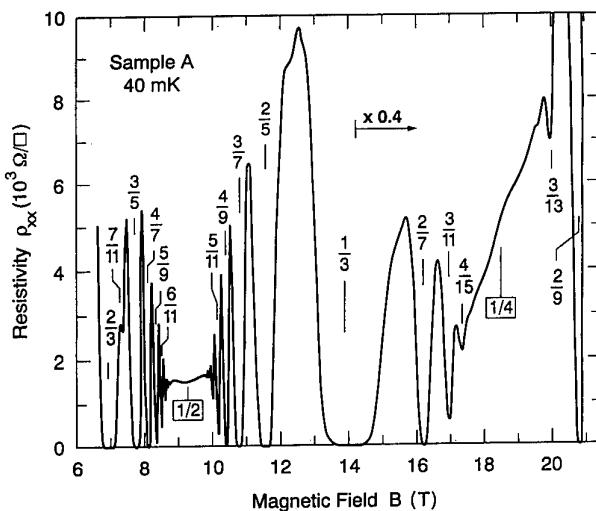


Figure 1. Overview of diagonal resistivity ρ_{xx} in the vicinity of $v = 1/2$ and $v = 1/4$ at $T = 40 \text{ mK}$. Landau level filling fractions are indicated. For fields higher than 14 T the data are divided by a factor of 2.5.

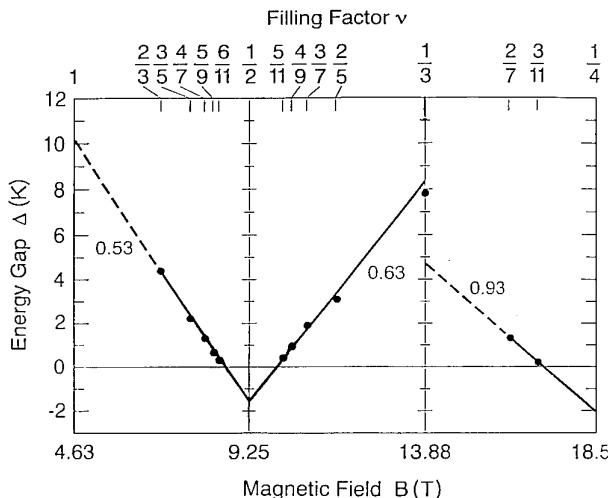


Figure 2. Gap energies for various filling factors in the vicinity of $v = 1/2$ and $v = 1/4$ in the sample of figure 1 plotted versus magnetic field. Straight lines are a guide to the eye. The number associated with each line represents the effective mass in units of m_e .

shows all the data in a single graph including two data points at $v = 2/7$ and $3/11$. Plotted in this fashion we observe a striking linearity between gap energies and magnetic field, reminiscent of the opening of the cyclotron energy gap for electrons around $B = 0$. One can characterize our findings phenomenologically in very simple terms once one establishes $B_{1/2}$ at $v = 1/2$ (and $B_{1/4}$ at $v = 1/4$) as a new origin for B and adopts its deviation B_{eff} from this value as the *effective magnetic field*. Just like in a simple fermion system of density n , the minima in ρ_{xx} due to the primary sequences of the FQHE occur at

$$\begin{aligned} \Delta B_p &= B_{p/(2p \pm 1)} - B_{1/2} = (\hbar n/e)[(2p \pm 1)/p - 2] \\ &= \pm \hbar n/e p. \end{aligned}$$

Since the gap energies are linear in B_{eff} they can be characterized by an effective mass m^* via $\hbar\omega_c = \hbar e B_{\text{eff}}/m^*$, assuming e to be the electronic charge. The gaps are reduced by a broadening Γ of the levels (taken to be field independent) which accounts for the absence of the highest-denominator energy gaps and for the negative intercept at $B_{\text{eff}} = 0$. The values for m^* (in units of the free electron mass, m_e) are indicated in figure 2. They are about one order of magnitude bigger than the band electron mass $m_b \simeq 0.07 m_e$ of GaAs.

These observations find a consistent interpretation in terms of the composite fermion model for higher-order fractions by Jain [15] and the recent proposal of the existence of a Fermi surface at $v = 1/2$ by HLR [17]. In this model, the primary sequences of FQHE states at $v = p/(2p \pm 1)$ around $v = 1/2$ become the Shubnikov-de Haas (SdH) oscillations of composite fermions in the presence of the effective magnetic field. The magnetic-field-dependent energy gap represents the Landau level splitting $\hbar\omega_c = \hbar e B_{\text{eff}}/m^*$ of the composites of charge e and mass m^* . Just as in the case of regular fermions, the negative intercept at $v = 1/2$ can then be interpreted as

the broadening Γ of the composite fermion ‘Landau levels’ which causes disappearance of the gap as $\Gamma \approx \hbar\omega_c$. A comparison of transport-scattering time and relaxation time around $B = 0$ and $B_{\text{eff}} = 0$ further strengthens the case for the existence of new particles [18] at $v = 1/2$.

The magnetic field dependence of experimental gap energies of the sequence of higher-order FQHE states around filling factor $v = 1/2$ find a remarkably consistent interpretation in terms of novel composite fermions which, driven by Coulomb interaction, coagulate from electrons and flux-quanta.

3. Shubnikov-de Haas oscillations

The analogy with regular SdH oscillations can be carried further [19] and the magnetoresistance data due to the FQHE around $v = 1/2$ can be interpreted in terms of the standard SdH formalism [20, 21] commonly used for electrons. It is important to differentiate this approach from the activation energy analysis [18] of the previous section.

The conceptual background for activation energy measurements is the traditional quantum liquid picture in which thermally activated fractionally charged quasi-particles are generated across an energy gap [4]. Likewise, the experimental procedure follows the traditional $\log \rho_{xx}$ versus $1/T$ data reduction of the FQHE, where ρ_{xx} is the value of the resistivity at the minima. By employing the standard SdH formalism to the oscillations around $v = 1/2$, one implicitly implies a Fermi sea with a fixed Fermi energy E_f and an initially smooth density of states (dos). The application of a magnetic field introduces periodic oscillations in this dos whose amplitude at E_f is probed by magnetotransport. Figure 3 shows magnetotransport data around $v = 1/2$ for four different temperatures. Already in this overview the oscillations seem to display the familiar behaviour of regular SdH oscillations around zero magnetic field. Apart from small distortions the data exhibit a line of mirror symmetry at $v = 1/2$, the oscillations have a finite onset about 0.7 T to either side of $v = 1/2$, their spacing is proportional to $1/B$ as measured from $v = 1/2$, and their amplitude shrinks with rising temperature.

The temperature dependence of the amplitude Δ_v follows extraordinarily well [19] the traditional sinh dependence of the SdH formalism usually denoted as

$$\Delta R/R_0 = (A_T/\sinh A_T)[4 \exp(-\pi/\omega_c \tau)] \quad (1)$$

with

$$A_T = 2\pi^2 k T / \hbar \omega_c$$

once the magnetic field is replaced by B_{eff} . The scattering rate $1/\tau$ is often expressed in terms of the Dingle temperature [22], T_D , usually defined as $kT_D = \hbar/\pi\tau$. The good agreement (not shown) between the experimental data and equation (1) gives one confidence in the applicability of the SdH formalism to the magnetoresistance arising from the FQHE states around $v = 1/2$.

The effective-mass values and their uncertainties deduced in this fashion from our data are shown in the

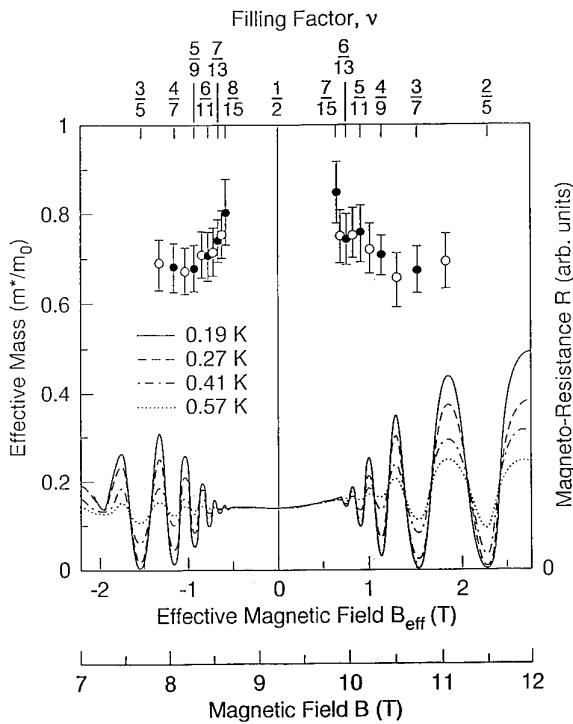


Figure 3. Temperature dependence of the magnetoresistance, R (bottom), and deduced effective masses, m^* (in units of bare electron mass m_e , top) as a function of magnetic field, B , and effective magnetic field, B_{eff} , around $v = 1/2$ Landau level filling factor.

upper part of figure 3. Over the range of magnetic fields which we can explore, and to within the accuracy of our data, the effective mass is constant, having a value of $\sim 0.7m_e$, in units of the free electron mass with possibly a slight upturn towards $v = 1/2$. The apparent increase of the mass as $B_{\text{eff}} \rightarrow 0$ is an interesting feature of the data. The theory by HLR actually proposes a mass enhancement in the vicinity of $v = 1/2$ due to interactions with a fluctuating gauge field which, for the case of Coulomb interactions, takes on the form of a logarithmic correction ([17], equation 6.44). However, since the mass increase for $v \rightarrow 1/2$ in figure 3 is comparable to the error bars and appears only in the weakest SdH oscillations, our data are insufficient to support experimentally such an asymptotic behaviour around $v = 1/2$.

The sum of our findings re-emphasizes the internal consistency of the SdH formalism when applied to the FQHE data. The deduced masses are similar to, although somewhat higher than, those observed earlier in thermal activation measurements [18] on the same specimen. This is largely due to different data reduction procedures as well as to the limited range explored in the SdH analysis, which excludes the lower-denominator fractions at large $|B_{\text{eff}}|$ due to interferences from the states at 3/8 and 5/8. A possible field dependence within the range of B_{eff} to which the SdH analysis has been applied is beyond the accuracy of our measurements, but has been proposed based on a similar SdH analysis of the FQHE states by Leadley *et al* [23]. Using an effective mass of $m^* = 0.7m_e$, one can deduce with the help of equation (1) the carrier

scattering time $\tau \sim 4.4(\pm 1) \times 10^{-12}$ s. This value for τ , equivalent to $\Gamma = \hbar/\tau = 1.7 \pm 0.4$ K, is similar to the value of $\Gamma = 1.4$ K derived for this quantity from the activation energy gap data [18].

In summary, the analysis of the resistance oscillations in the FQHE around $v = 1/2$, in terms of the well established SdH formalism, traditionally used for electrons, turns out to be very successful. The results of such a conventional analysis in terms of a simple, non-interacting carrier model, are internally consistent, yield effective masses and scattering rates, and further strengthen the case for the existence of a new particle around $v = 1/2$.

4. How real are composite fermions?

Although the above measurements provide considerable experimental support for the existence of composite fermions, it is natural to wonder just how real these particles are. It seems unsatisfactory to simply regard them as a convenient mathematical construct. In fact, it would be far more satisfying if one could detect a semiclassical aspect of the composites. Since the $v = 1/2$ state is proposed to be largely equivalent to a metal at zero magnetic field, one wonders whether experiments that usually reveal the semiclassical motion of electrons as a dimensional resonance could not be performed on these new particles. Electronic transport through antidot superlattices [24] yields particularly strong dimensional resonances. The resistivity of the patterned two-dimensional electron gas shows a sequence of strong peaks at low magnetic field. A simple geometrical construct reveals that the resonances occur when the classical cyclotron orbit, $r_c = m^*v_F/eB = \hbar k_F/eB$, (m^* is the effective mass, v_F is the Fermi velocity, $k_F = (2\pi n_e)^{1/2}$ is the Fermi wavevector and n_e is the electron density), encircles a specific number of antidots. The inset in figure 4 illustrates the configurations for $s = 1, 4$ and 9 dots. According to a simple electron ‘pinball’ model [24], at these magnetic fields the orbit is minimally scattered by the regular dot pattern, electrons get ‘pinned’ and transport across the sample is impeded. In a more sophisticated model the resistivity peaks arrive from the correlation of chaotic, classical trajectories [25].

Kang *et al* [26] used such an antidot superlattice to first establish the semiclassical behaviour of electrons around $B = 0$ and then probe the equivalent semiclassical behaviour of these bizarre composite particles around $v = 1/2$ in a sample of $n = 1.45 \times 10^{11} \text{ cm}^{-2}$ and $\mu = 7.8 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Figure 4 shows the magnetoresistance ρ_{xx} of a $d = 600 \text{ nm}$ period antidot superlattice in comparison with ρ_{xx} of the unprocessed bulk part of the sample, devoid of dots. Both specimens behave very similarly showing FQHE states as high as $v = 9/17$. However, a striking difference in ρ_{xx} is apparent near $B = 0$, $v = 3/2$ and $v = 1/2$. The bulk sample clearly exhibits local minima at these field positions, whereas the resistance in the antidot trace shows overall maxima with peaks of varying strength superimposed at the same fields. The features around $B = 0$ are clearly identifiable

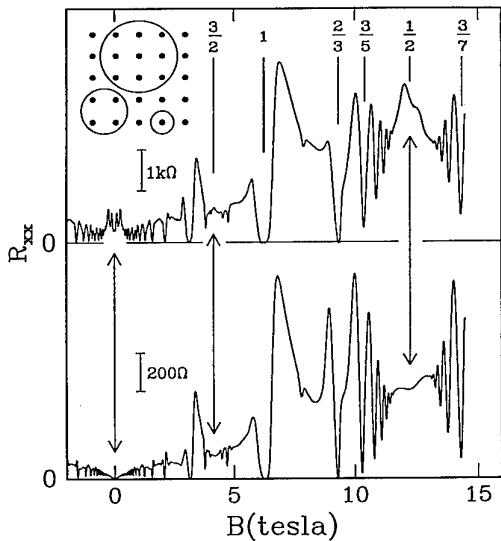


Figure 4. Comparison of the magnetoresistance ρ_{xx} of the bulk two-dimensional electron gas (lower trace) and ρ_{xx} of the $d = 600$ nm period antidot superlattice (upper trace) at $T = 300$ mK. The fractions near the top of the figure indicate the Landau level filling factor. Inset: schematic diagram of commensurate orbits encircling $s = 1, 4$ and 9 antidots.

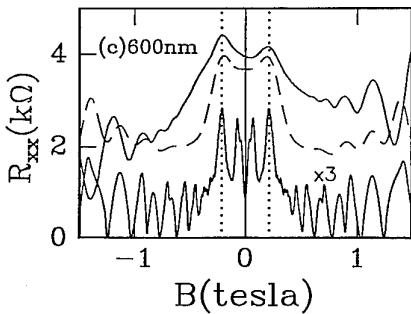


Figure 5. Expanded view of the magnetoresistance near $v = 1/2$ and $B = 0$ for a 600 nm period antidot superlattice. The $v = 1/2$ result is shown as the upper trace. It has been shifted to zero and the field scale has been divided by $\sqrt{2}$ for comparison. The vertical scale reflects the resistance for the $v = 1/2$ traces. The electron trace has been multiplied by 3. The broken curve shows simulated smearing by Fourier filtering of the electron trace.

as the well known dimensional resonances of the electrons. The peaks around $v = 1/2$ are the sought-after dimensional resonances of composite fermions. A direct comparison between electron and composite resonances is made in the following figure.

Figure 5 shows the ρ_{xx} data around $B = 0$ and $v = 1/2$ with an antidot superlattice of periodicity of 600 nm. The origin, $B = 0$, resides at the centre of the figure and the lower traces in each section represent the electron transport data taken for positive and negative magnetic fields. The well established electron dimensional resonances are clearly visible. The top trace represents ρ_{xx} around $v = 1/2$ shifted to $B = 0$. Concomitantly, the

magnetic field scale is compressed by a factor of $\sqrt{2}$ to account for the expected difference in k_F between electrons and composites due to the lifting of the spin degeneracy. In comparing the dimensional resonances around $B = 0$ with those around $v = 1/2$, we observe excellent agreement between the strong features of the electrons and the peaks in the respective traces at half-filling. This is compelling evidence that these magnetoresistance peaks around $v = 1/2$ are due to antidot dimensional resonances of composite fermions. Similar dimensional resonances of composite fermions have also been observed in surface acoustic wave experiments by Willett *et al* [27] and in magnetic focusing experiments by Goldman and Su [28].

The observation of dimensional resonances and their appropriate scaling demonstrates the semiclassical motion of composite fermions and suggests that in transport experiments around $v = 1/2$ these new particles, in many aspects, behave like ordinary electrons. It is remarkable that the complex electron-electron interaction in the presence of a magnetic field can be described in such simple, semiclassical terms.

The introduction of the composite fermion model has provided us with a new and powerful physical picture for the fractional quantum Hall effect.

References

- [1] Tsui D C, Stormer H L and Gossard A C 1982 *Phys. Rev. Lett.* **48** 1559
- [2] Prange R E and Girvin S M (ed) 1990 *The Quantum Hall Effect* (New York: Springer)
- [3] Chakraborty T and Pietilainen P 1988 *The Fractional Quantum Hall Effect* (New York: Springer)
- [4] Laughlin R B 1983 *Phys. Rev. Lett.* **50** 1395
- [5] Haldane F D 1983 *Phys. Rev. Lett.* **51** 605
- [6] Halperin B I 1984 *Phys. Rev. Lett.* **52** 1583; 1984 *Phys. Rev. Lett.* **52** 2390(E)
- [7] Laughlin R B 1990 *The Quantum Hall Effect* ed R E Prange and S M Girvin (New York: Springer) pp 233–302
- [8] Morf R, d'Ambrumenil N and Halperin B I 1986 *Phys. Rev. B* **34** 3037
- [9] Yoshioka D, MacDonald A H and Girvin S M 1988 *Phys. Rev. B* **34** 3636
- [10] Bérán P and Morf R 1991 *Phys. Rev. B* **43** 12654
- [11] Willett R L, Eisenstein J P, Stormer H L, Tsui D C, Gossard A C and English J H 1987 *Phys. Rev. Lett.* **59** 1776
- [12] Halperin B I 1983 *Helv. Phys. Acta* **56** 75
- [13] Jiang H W, Stormer H L, Tsui D C, Pfeiffer L N and West K W 1989 *Phys. Rev. B* **40** 12013; 1992 *Phys. Rev. B* **46** 10468
- [14] Willett R L, Paalanen M A, Ruel R R, West K W, Pfeiffer L N and Bishop D J 1990 *Phys. Rev. Lett.* **65** 112
- [15] Jain J K 1989 *Phys. Rev. Lett.* **63** 199; 1989 *Phys. Rev. B* **40** 8079; 1990 *Phys. Rev. B* **41** 7653
- [16] Dev G and Jain J K 1991 *Phys. Rev. B* **45** 1223
- [17] Halperin B I, Lee P A and Read N 1993 *Phys. Rev. B* **47** 7312
- [18] Du R R, Stormer H L, Tsui D C, Pfeiffer L N and West K W 1993 *Phys. Rev. Lett.* **70** 2944
- [19] Du R R, Stormer H L, Tsui D C, Pfeiffer L N and West K W 1994 *Solid State Commun.* **90** 71

- [20] Shubnikov L and de Haas W J 1930 *Leiden Commun.* 207a, 207c, 207d, 210a
- [21] Adams E N and Holstein T D 1959 *J. Phys. Chem. Solids* **10** 254
- [22] Dingle R B 1952 *Proc. R. Soc. A* **211** 517
- [23] Leadley D R, Nicholas R J, Foxon C T and Harris J J 1994 *Phys. Rev. Lett.* **72** 1906
- [24] Weiss D, Roukes M L, Menschig A, Grambow P, von Klitzing K and Weimann G 1991 *Phys. Rev. Lett.* **66** 2790
- [25] Fleischmann R, Geisel T and Ketzmerick R 1992 *Phys. Rev. Lett.* **68** 1367
- [26] Kang W, Stormer H L, Pfeiffer L N, Baldwin K W and West K W 1993 *Phys. Rev. Lett.* **71** 3850
- [27] Willett R L, Ruel R R, West K W and Pfeiffer L N 1993 *Phys. Rev. Lett.* **71** 3846
- [28] Goldman V J, Su B and Jain J K 1994 *Phys. Rev. Lett.* **72** 2065

Theory of the half-filled Landau level

N Read

Departments of Physics and Applied Physics, PO Box 208284, Yale University,
New Haven, CT 06520, USA

Abstract. A recent theory of a compressible Fermi-liquid-like state at Landau level filling factors $v = 1/q$ or $1 - 1/q$, q even, is reviewed, with emphasis on the basic physical concepts.

The physics of interacting electrons in two dimensions in very high magnetic fields has proved to be a rich subject since the discovery of the FQHE (fractional quantum Hall effect) in 1982 [1]. The initial motivation for study in this area was the expectation that a Wigner crystal should form when the (areal) density of electrons is low and the magnetic field is high, not only because, as at low magnetic field, the Coulomb energy of repulsion dominates the kinetic energy at low density, but also because if the magnetic field is high enough so that mixing of excitations to higher Landau levels can be neglected then the kinetic energy itself becomes essentially constant. In this limit the electrons would behave classically, with their dynamics given by the $E \times B$ drift of their guiding centre coordinates, i.e. drift along the equipotentials, the potential being here given by the Coulomb potential of the other electrons. The Wigner crystal is presumably the unique lowest-energy, stationary state at a given density in this classical problem. The surprise was that in fact, at accessible densities, the quantum fluctuations cause this crystal to melt (as parameters are varied) into some quantum fluid, and furthermore the nature of this fluid depends on the commensuration between the density and the magnetic field, in an essentially quantum mechanical way. In terms of the Landau level filling factor $v \equiv n\Phi_0/B$ (where n is the density, B is the magnetic field strength and $\Phi_0 = hc/e$ is the flux quantum), this was manifested by quantum Hall effect (QHE) plateaus at $v = p/q$ a rational number with an odd denominator q (p and q have no common factors) in all except a handful of special cases observed more recently.

A theory that explained many features of this picture, including the ‘odd denominator rule’, was quickly forthcoming [2–4]. Laughlin’s states at $v = 1/q$, q odd, and the hierarchical extension to all $v = p/q$, q odd, are incompressible fluids capable of exhibiting a QHE plateau at $\sigma_{xy} = (p/q)e^2/h$. They also possess novel elementary excitations called quasiparticles, which carry a fraction $\pm 1/q$ of the charge on an electron, and have fractional statistics, that is the phase of the wavefunction changes by $e^{i\theta}, \theta/\pi = \pm p'/q$ when two quasiparticles are exchanged adiabatically [4–7]. A finite energy Δ is required to create one of these excitations. The fractional QHE associated

with a particular state will be destroyed if the energy scale characterizing the RMS disorder exceeds about Δ .

Some questions remained unanswered, however. While the hierarchy theory predicted that states at v with larger denominators would have smaller gaps, it would seem, at least naively, that all fractions with the same denominator would be equally strong, though it is possible that those near a stronger, smaller-denominator state might be overwhelmed, for example $3/11$ near $1/3$. As samples improved, it could be seen that in fact most fractions observed were of the restricted form $v = p/(mp + 1)$, $p \neq 0$, $m \geq 0$ even, $= 2, 4$. It was not clear if the hierarchy could explain this, although it could be simply a quantitative fact, without deeper explanation. Furthermore, in the region near $v = 1/2$, where $m = 2$, p is large in the above formula, no fractional plateaus were seen, but there was a shallow minimum in ρ_{xx} that remained $\neq 0$ as $T \rightarrow 0$. A possibly more profound theoretical question was, what is the nature of the ground state at fillings not of the hierarchy form p/q , q odd? Indeed, the construction of states at $v = p/q$, q odd that exhibit FQHE does not prove that FQHE cannot occur at even q . The whole question of what *does* occur received almost no attention in published work until the last few years (a notable exception is the FQHE at $v = 5/2$ and the Haldane–Rezayi proposal for its explanation [8, 9]). One possibility is that there is no well defined state at these fillings in the thermodynamic limit, that phase separation into domains of two nearby, stable FQHE states occurs instead. Another possibility is a well defined pure phase constructed by taking a FQHE state at a nearby filling, and adding a low density of, say, quasiholes which at low density could form a Wigner crystal. Such a state could exist (with varying lattice constant) over a range of densities. These possibilities may actually occur near some particularly stable (large Δ) fractions such as $1/3$, but seem less likely near $v = 1/2$ since neighbouring states would have very small energy gaps. If liquid states exist at non-FQHE filling factors, then they lie outside the then-current theoretical understanding.

With these motivational remarks, we now abandon the historical approach in order to develop the overall theoretical picture [10–12] as logically as possible. Two

alternative approaches will be described. The first begins close to Laughlin's original ideas, and contains explicit trial wavefunctions, as well as other notions of the last few years. The other approach is field theoretical and eventually arrives at the same conclusions, but may seem less physical at first. However, it is much more appealing for explicit analytical calculation. The relation with Jain's work on the hierarchy states will also be explained.

Let us turn, then, to the first of these approaches. We make the usual assumption that, at $T = 0$, interaction energies $\sim v^{1/2} e^2 / \epsilon l$ are weak compared with the Landau level splitting $\hbar\omega_c$, and so the electrons should all be in the lowest Landau level (LLL) with spins polarized when $v < 1$. We work in the plane with complex coordinate $z_j = x_j + iy_j$ for the j th electron. In the symmetric gauge [2], single-particle wavefunctions in the LLL are $z^m \exp(-\frac{1}{4}|z|^2)$, where we set $\hbar = 1$, and the magnetic length $l = \sqrt{\hbar c/eB} = 1$. m is the angular momentum of the state. An N -particle wavefunction has the form

$$\Psi(z_1, \dots, z_N, \bar{z}_1, \dots, \bar{z}_N) = f(z_1, \dots, z_N) \exp\left(-\frac{1}{4} \sum_i |z_i|^2\right) \quad (1)$$

where f is complex analytic and totally antisymmetric. If f is homogeneous of degree M in each z_i , it has definite total angular momentum and describes a (not necessarily uniform) droplet of radius $\sim \sqrt{2M}$. As a function of each z_i , f has M zeros which is also the number of flux quanta N_ϕ enclosed by the circle of radius $\sqrt{2M}$. (Similarly, on a closed surface such as the sphere, the number of zeros of the LLL single-particle wavefunctions equals the number of flux quanta through the surface.) Thus the filling factor $v = N/N_\phi$ = the number of particles per flux quantum.

Now if f contains a factor

$$U(z) = \prod_i (z_i - z) \quad (2)$$

then one zero for each particle is located at z . ($U(z)$ can be viewed as an operator, Laughlin's quasi-hole operator, acting on the remainder of the wavefunction by multiplication to produce f .) This means that there are no particles in the immediate vicinity of z , there is a depletion of charge there. We describe this as a *vortex* at z , since as in a vortex in a superfluid, the phase of the wavefunction winds by 2π as any z_i makes a circuit around z .

Given a state, we can obtain a valid new state by multiplying by $U(z)$. (This increases M by 1 if $z = 0$, but otherwise by an indefinite amount. But it always increases N_ϕ by 1 if N_ϕ is defined as the flux through the region occupied by the fluid.) For reasonably uniform fluid states (such as Laughlin's, or the Fermi liquid below) the vortex can be considered as an excitation of the fluid. Now add in addition an electron. Clearly it is attracted to the centre of the vortex, due to the density deficiency there. Since there is no kinetic energy, it can certainly form a bound state. Similarly it can bind to multiple vortices $U(z)^q$.

It is natural to consider the possibility that the ground

state itself contains electrons bound to vortices, since this will give a low energy. As each vortex is added, N_ϕ increases by 1. If we add one electron for every q vortices, we can form identical bound states, and if all zeros are introduced in this way, we will have $N_\phi = q(N - 1)$, which implies $v = N/N_\phi \rightarrow 1/q$ as $N \rightarrow \infty$. This case is by far the simplest to understand. In this state, there will be q -fold zeros as one electron approaches another. It has long been realized that Laughlin's Jastrow-like ansatz

$$f_L = \prod_{i < j} (z_i - z_j)^q \quad (3)$$

can be called a 'binding of zeros to particles'.

However, the astute reader will already be aware that f_L is antisymmetric only for q odd. To solve this problem, and to obtain a deeper insight into Laughlin's state, we must examine the nature of the bound objects more closely. Since we know the properties of an electron, we turn to the vortices. First, the q -fold vortex carries a charge $-vq$ in a fluid state at arbitrary filling factor v (this includes our compressible Fermi liquid state below, as well as the usual hierarchy states). This can be established by Laughlin's plasma analogy [2] or indirectly through adiabatic motion of the vortex [5]; the latter will be more useful here. A crude version of this argument is that when a vortex is moved around adiabatically in some given fluid state, the wavefunction picks up a phase, since the definition (2) of $U(z)$ shows that it changes phase by 2π for each particle about which it makes a circuit. Then if it makes a circuit around a (non-self-intersecting) closed loop enclosing a region D of area A , it will pick up a phase $2\pi \int_D d^2z \rho(z)$ which reduces to vA if the density $\rho(z)$ is uniform $\rho = v/2\pi$. This is then identified as the same result as would be obtained for a particle of charge $-v$ in the magnetic field seen by the electrons. Similarly, a q -fold vortex has a charge deficiency of vq , which for $v = 1/q$ is equivalent to a real hole, so the electron- q -vortex composite at this filling has net charge zero, and behaves like a particle in zero magnetic field. But note that the vortex is actually sensitive to the density of electrons, which can vary in space and time, even when the external magnetic field and the average filling factor are fixed.

Now for the famous fractional statistics of vortices. If a vortex makes a circuit around a loop enclosing another vortex, with the density otherwise uniform, the missing charge around the vortex core will make a difference of $2\pi v$ to the phase (independent of the size and shape of the loop, as long as the vortices remain far enough apart). But a circuit is equivalent to two exchanges, up to translations. So a similar calculation gives that adiabatic exchange of two vortices produces a phase $\theta = \pi v$. For two q -fold vortices, the result is $\pi q^2 v$ which at $v = 1/q$ is πq . This shows that at these fillings, q -fold vortices are fermions for q odd, and bosons for q even. Hence the electron- q -vortex bound state is a boson for q odd, and a fermion for q even. For q odd, we can now argue that the bosons in zero magnetic field can Bose condense (at $T = 0$) into the zero-momentum state

and that this is the interpretation of the Laughlin state [13]. If ψ^\dagger creates an electron in the LLL, then $\psi^\dagger(z)U^q(z)$ creates a boson, and the condensate is obtained by letting $\int d^2z\psi^\dagger(z)U^q(z)\exp(-\frac{1}{4}|z|^2)$ act repeatedly on the vacuum. This produces exactly the Laughlin state [13]. Note that condensation arises because particles try to minimize their kinetic energy $\sim k^2$. We need to show that the bound objects do in fact have such an *effective* ‘kinetic’ energy. The true kinetic energy of the electrons has been quenched, so this term in the effective Hamiltonian for the bound objects can only arise from electron–electron interactions. Before discussing how this arises, let us pursue the consequences. At $v = 1/q$, q even, the bound objects are fermions in zero net field, which cannot Bose condense but which can form a Fermi sea. This is then the proposal for a compressible state at these fillings. It will be compressible because it can be shown that incompressibility, and the quantized Hall effect, in the case of bosonic bound states is a consequence of Bose condensation (superfluidity), which does not occur in a Fermi sea unless BCS pairing occurs, in which case the state becomes a QHE state. The Haldane–Rezayi and Pfaffian states arise in this way [14].

We now return to the dynamics of the bound objects, or ‘quasiparticles’. We have argued that there is an attraction between an electron and a q -fold vortex. In a uniform background, this will take the form of a central potential. The minimum is at the centre, meaning that the electron is exactly at the zeros of the many-particle wavefunction. Now at filling factor $1/q$, we need to consider quasiparticles in plane wave states with wavevector \mathbf{k} . These can be created by acting on a fluid with $\int d^2z e^{ik \cdot r} \psi^\dagger(z)U^q(z)\exp(-\frac{1}{4}|z|^2)$. In the wavefunction, this means a factor $\exp(i\mathbf{k} \cdot \mathbf{r}_j)$ in the term where the j th electron is the bound electron in the state with wavevector \mathbf{k} . Now it turns out that this factor, acting by multiplication on some given state, displaces the j th particle by $i\mathbf{k}$, where $\mathbf{k} = k_x + ik_y$ (recall that the magnetic length l is 1). This arises because in the Hilbert space of many-particle states of the form (1), \bar{z}_j acts on f as $2\partial/\partial z_j$, which generates displacements [15]. A quasiparticle with wavevector zero would have the electron exactly at the zeros of the wavefunction. So a quasiparticle with wavevector \mathbf{k} has the electron displaced by $|\mathbf{k}|$ from the centre of the vortex. The electron and vortex experience a potential $V(|\mathbf{k}|)$ due to the Coulomb repulsion of the electron by the other electrons, which are excluded from the vortex core. A good understanding is now achieved semiclassically. The electron will drift along an equipotential of $V(|\mathbf{k}|)$. From the preceding discussion, at $v = 1/q$ the q -fold vortex experiences a magnetic field of the same strength as the electron, and so it will also drift with the same speed but in the opposite sense relative to the gradient of the potential, due to the opposite sign of its effective charge. This means that both components of the pair drift in the same direction, perpendicular to the vector connecting their centres, so that their separation ($=|\mathbf{k}|$) remains constant. The picture is like that of oppositely charged particles in a magnetic field, which can drift in a straight line as a pair. The energy of our

pair is $V(|\mathbf{k}|)$ and the velocity is $\propto \partial V / \partial |\mathbf{k}|$ as it should be. Near the bottom of the potential it will be quadratic, and we can obtain an effective mass $\sim (\partial^2 V / \partial |\mathbf{k}|^2)^{-1}$ due to the interactions (a similar calculation was performed in [13]). This shows clearly that the interactions favour condensation of the quasiparticles to minimize this effective kinetic energy. In the q even case, the quasiparticles are fermions and must have distinct \mathbf{k} s, filling a Fermi sea. Then in the ground state not all the zeros of the wavefunction are precisely on the electrons but some are displaced by amounts up to k_F , determined in the usual way by the density, $k_F^2/4\pi = 1/2\pi q$. For q large these displacements $\sim q^{-1/2}$ are small compared with the interparticle spacing $\sim q^{1/2}$. Thus a trial state of this form may not be much worse energetically than the Laughlin state at nearby odd q . We believe that this is the essential reason why this idea has a good chance of being the correct many-body ground state: it is due to the good correlations that produce a low Coulomb energy. Of course, if q is very small, this picture might break down, but in fact we have good reason to believe it holds for q as small as 2. This argument also tells us that low-lying excited states are obtained by increasing the wavevector of a quasiparticle in the Fermi sea. For q large, the Fermi velocity will be determined by the same effective mass as near the bottom of the sea; otherwise we must take the derivative at k_F .

To consider collective effects it is necessary to go beyond an independent quasiparticle picture. There are important long-range interactions that can be described as gauge fields. We will examine these in the context of the field theoretic approach later. As a test of the above ideas, we can perform numerical diagonalization of small systems at, say, $v = 1/2$. This has been done recently in a paper by Rezayi and Read [12]. Excellent agreement of trial states suggested by the above ideas is found with the exact wavefunctions of low-lying states. The trial wavefunctions were generated on the sphere, but on the plane would be roughly

$$\Psi = \mathcal{P}_{\text{LLL}} \det \mathbf{M} \prod_{i < j} (z_i - z_j)^q \exp\left(-\frac{1}{4} \sum |z_i|^2\right). \quad (4)$$

The matrix \mathbf{M} has elements that are essentially plane waves $M_{ij} \sim \exp(i\mathbf{k}_i \cdot \mathbf{r}_j)$ for the quasiparticles, filling the Fermi sea. \mathcal{P}_{LLL} projects all electrons to the LLL. As explained above, the plane wave factors then act as operators within the LLL, on the Jastrow factor $\prod (z_i - z_j)^q$ which if not modified would be the Laughlin state for bosons at $v = 1/q$. The Slater determinant makes the state totally antisymmetric. The simple product form is similar to Jain’s states off half-filling [11]; it differs a little, but insignificantly, from the present author’s first idea (in 1987) of building the Fermi sea by acting with Fourier components of $\psi^\dagger(z)U^q(z)\exp(-\frac{1}{4}|z|^2)$, which was suggested by the analogy with the Laughlin states at q odd [13]. The numerical calculations also showed that the two-particle correlation function $g(r)$ in the ground state possesses (i) a large ‘correlation hole’ at short distances r , consistent with the argument that zeros are bound close to the electrons, and (ii) oscillations at large r , perhaps

of the form $r^{-\alpha} \sin 2k_F r$ asymptotically, as in a two-dimensional Fermi gas, which has $\alpha = 3$.

We now turn to the field theoretic approach. It begins with the observation that in two dimensions particles of any (fractional) statistics can be represented by charged particles of other statistics attached to δ -function flux tubes of a certain size [16]. A charged particle dragged adiabatically around a flux tube (or vice versa) picks up an Aharonov–Bohm phase of 2π times the product of charge of the particle and the number of flux quanta in the tube. Thus if identical particles are attached to identical flux tubes, a circuit of one composite around the other produces a phase 4π times the charge times the flux, since each particle sees the other flux. For an exchange, we get only half this. In addition, the exchange of identical particles produces a phase $e^{i\theta}$ due to the statistics of the particles themselves. Thus fermions or bosons attached to fractional flux tubes can be used to model anyons, where the total phase obtained in an exchange is fractional [16]. In the fractional quantum Hall effect we change our terminology slightly because we view flux tubes as operators that act only on particles already present. Thus a composite is introduced by adding first a flux tube, then a particle is added to the same point. The phase produced by an exchange is then only π times the charge times the flux in each composite, plus the phase due to exchanging the particles themselves. Thus, for example, we can say that attaching two flux quanta to each boson in a system leaves the composites still as bosons, or doing the same to fermions leaves them fermions. Then in the high-field situation of interest here, we may represent electrons as q (integer) flux tubes attached to some other particles. The latter must be bosons if q is odd, and fermions if q is even, in order to reproduce the Fermi statistics of the electrons. The flux tubes can be represented by a ‘fictitious’ vector potential \mathbf{a} , not to be confused with the physical vector potential \mathbf{A} representing the constant external field, which obeys

$$\nabla \times \mathbf{a} = 2\pi q\rho \quad (5)$$

and the density $\rho(\mathbf{r}) = \sum \delta^{(2)}(\mathbf{r}_i - \mathbf{r})$.

The next step, used in several similar problems [10, 11, 17, 18], is a mean field approximation. The new fermions (for q even) see the constant background magnetic field, and the q flux tubes attached to the other fermions. If the quantum mechanical state has a uniform density (in the quantum average), the latter becomes a constant field whose sign we can choose to be opposite to the external field. In particular, if $v = 1/q$ the fields cancel exactly. The fermions now see no net field, so they can form a Fermi sea, which does have a uniform density, as we assumed. We have arrived at the physical picture of the compressible Fermi-liquid-like state, where we can say loosely that the fermions are electrons plus q flux quanta. But notice that this should not be taken too literally, since we have actually just made a transformation. The real flux due to the external field remains uniform, not bunched up into flux tubes attached to the electrons. It is really vorticity that is bound to the

electrons, as discussed above. In mean field theory the wavefunction for the electrons is simply

$$\Psi_{MF} = \det \mathbf{M} \prod_{i < j} (z_i - z_j)^q / |z_i - z_j|^q \quad (6)$$

with \mathbf{M} as before and no projection. This function is *not* all in the LLL. We can see that the effect of the transformation and the mean field approximation is to build the right kind of phase factors into the wavefunction, the same as possessed by q vortices on each electron. The factor $\prod |z_i - z_j|^q$ needed to recover equation (4) can in fact be obtained from fluctuations about mean field, at least in a long-wavelength sense [19].

Excited states again involve creation of particle–hole pairs. In the mean field approximation, the effective mass of fermion excitations near the Fermi surface is simply the bare electron mass m , since that is what appears in the Hamiltonian (we do not attempt to impose the LLL constraint). Consideration of collective oscillations of the system leads to the correct cyclotron frequency $\omega_c = eB/mc$ which according to Kohn’s theorem cannot be renormalized. However, under the conditions stated at the beginning of this paper, the low-energy Fermi excitations should have an effective mass m^* determined by interelectron interactions only. Part of the resolution of this problem is that if fluctuations renormalize the effective mass, as they usually do in Fermi liquids, then there will also be a Landau interaction parameter F_1 that obeys a relation $m^{-1} = m^{*-1} + F_1$ which guarantees that the collective mode frequency is unrenormalized (like the plasma frequency which is unrenormalized in the usual electron gas at zero magnetic field) [10]. A more serious problem is that interactions seem to play no role in the mean field theory. The prediction would be the same, even for non-interacting electrons. The Fermi liquid mean field state has finite compressibility, due to the mass m , even though in this case a partially filled Landau level should have infinite compressibility! (The same holds when this approach is applied to the FQHE states for q odd, where the compressibility vanishes.) These problems can be resolved by understanding in what limit the approach is valid. Mean field theory is good, and fluctuations in the effective magnetic field $\nabla \times (\mathbf{A} + \mathbf{a})$ are small, when q is small. But for us q is an even integer ≥ 2 . However, if we replace the original problem of electrons in a magnetic field by that of anyons in a magnetic field, then we can make q small. We choose to study anyons with statistics $\theta = \pi(1+q)(\text{mod } 2\pi)$ and filling factor $1/q$, for $q > 0$. As $q \rightarrow 0$, the magnetic field goes to zero, but for all q we can still map the problem of anyons to one of fermions in zero net average magnetic field. So at $q = 0$ we reach fermions at zero magnetic field and we attach zero flux to convert them to fermions! At this point there can be no fluctuations in the effective magnetic field, since there is no attached flux. As we increase the external field, we must attach more flux to map to fermions in zero field, so the statistics must change accordingly. When θ has made a full circle from π back to π , we recover fermions (electrons) but now at a high field and filling factor $1/q$, $q = 2$. We can continue

and reach other even denominator states for electrons. This idea is an adaptation of that in [20], except that they use it to argue that certain states exist by adiabatic continuation, while we use it here only for a perturbation expansion in the fluctuations, i.e. in powers of q . Notice that in the limit $q = 0$, the system is a Fermi liquid, and the effective mass may be close to the bare mass if interactions are weak compared with the Fermi energy. A similar statement applies to the anyons for q small. It is only at $q = 2$ that we can argue that the compressibility must be infinite when interactions are set to zero, because only for fermions (such as the electrons) or charged bosons, *and not for anyons*, can we argue that the many-particle states of a non-interacting system are (anti-)symmetrized products of Landau level states.

After that rather technical paragraph, we return to simpler discussion. If the filling factor for electrons deviates from $v = 1/q$, q even, then in either approach above the quasiparticles will see a non-zero net magnetic field $B_{\text{eff}} = B - B_{1/2} = \nabla \times (\mathbf{A} + \mathbf{a})$. If the quasiparticles fill an integer number p of Landau levels due to this effective field, there will be a gap in the excitation spectrum (at least in mean field theory, but in fact it survives fluctuations) and we will obtain an integer QHE for the fermions, which is an FQHE for the electrons, with $\sigma = (e^2/h)p/(qp + 1)$. This of course is Jain's picture [11] of the 'main sequence' of fractional QHE states as observed in experiments. Jain's approach was a hybrid of the approaches above. He begins with the transformation involving δ -function fluxes, but then simply modifies the mean field wavefunctions (6) into the form (4) with M_{ij} = Landau level wavefunctions. The argument of binding zeros to particles to obtain a low Coulomb energy is not used. We believe the $v = 1/q$ case with exactly q vortices per particle gives valuable insight into the reasons for the stability of the states, even away from this filling. Also the notion of the effective mass emerges clearly in the Fermi liquid state. We predicted [10] that energy gaps in FQHE states near half-filling should scale as $1/p$ as $p \rightarrow \infty$, since the gap can be interpreted as an effective cyclotron energy $\sim B_{\text{eff}}/m^*$ once we recognize the existence of the Fermi liquid, with a well defined effective mass that controls excitation energies, as the limiting behaviour. This prediction has received experimental support [21].

The above arguments about dynamics of the quasi-particles at $v = 1/2$ also generalizes to this case. While the electron and the q -fold vortex still see the same potential $V(|k|)$, the drift velocities are different because that for the vortices is fixed by the electron density, not the true magnetic field. The separation still remains constant, so like the back wheels of a car turning a corner, the bound pair moves (semiclassically) on two concentric circles separated by $|k|$, giving an effective cyclotron radius for the quasiparticles. This radius is given by the usual formula $\Phi_0 k / 2\pi B_{\text{eff}}$ for a particle of charge 1 in a magnetic field B_{eff} .

The reader is cautioned that away from half-filling the quasiparticles which seem to be fermions in the mean field approach in fact have their statistics modified

because of screening effects on the fluctuations. They also acquire fractional charge, whereas at half-filling there is perfect screening by the response of the fluid to the δ -function of flux on the fermion. This effect is in fact described by the extra amplitude factors present in the wavefunctions (4). The arguments given above for electron- q -vortex bound states at $v = 1/q$ extend to other fillings to predict a net charge $1 - qv$ and statistics $\theta = \pi(q^2v - 1)$. These properties of fermions excited to higher 'quasiparticle Landau levels' at Jain's fractions are identical to those of the quasiparticle excitations in the hierarchy scheme, which is one of the main pieces of evidence for the equivalence of these two approaches [6, 7].

Space allows only a brief further discussion. An outstanding series of experiments has been performed by Willett and co-workers [22]. Their technique, using the propagation of surface acoustic waves across the surface of the device, probes the longitudinal conductivity of the two-dimensional electron gas at finite wavevectors and frequencies (this and other experiments [23] are discussed in the paper by Störmer in this issue). Since the velocity of the waves is slow compared with the Fermi velocity of our fluid, we can consider zero-frequency, finite wavevector. The theory, treating the response of the fermionic quasiparticles in an RPA-like approximation, shows that $\sigma_{xx}(k)$ increases linearly with wavevector k in the compressible state, as observed. This results from the transverse conductivity $\sim 1/k$ for a conventional Fermi liquid, on including the long-range effects of the gauge field \mathbf{a} . (Not only can $B_{\text{eff}} = 2\pi q[n - (2\pi q)^{-1}]$ fluctuate in space and time, but so can $E_{\text{eff}} = 2\pi qJ$, the effective electric field due to the current J in the perpendicular direction.) It also predicts resonances in the response when the wavevector is an integer multiple of the inverse cyclotron radius for the semiclassical motion of the quasi-particles close to but just off half-filling. This provides a measure of the Fermi wavevector. The recent experimental observation of these resonances confirms the (non-trivial!) existence of a Fermi surface for the charge-carrying excitations and excellent agreement of k_F with the expected value ($\sqrt{2}$ times that in zero magnetic field, because of spin) is obtained.

Open questions: a direct measure of the effective mass m^* at half-filling would be most welcome. Theoretically, a controversy remains about the possible partial breakdown of Fermi liquid theory, including behaviour of m^* , due to the fluctuations of \mathbf{a} [10].

Acknowledgments

I am grateful to P A Lee, B I Halperin and E Rezayi for collaborations. This work was supported by NSF-DMR-9157484 and by the A P Sloan Foundation.

References

[1] Tsui D C, Störmer H and Gossard A 1982 *Phys. Rev. Lett.* **48** 1559

- [2] Laughlin R B 1983 *Phys. Rev. Lett.* **50** 1395
- [3] Haldane F D M 1983 *Phys. Rev. Lett.* **51** 605
- [4] Halperin B I 1984 *Phys. Rev. Lett.* **52** 1583
- [5] Arovas D, Schrieffer J R and Wilczek F 1984 *Phys. Rev. Lett.* **53** 722
- [6] Read N 1990 *Phys. Rev. Lett.* **65** 1502
- [7] Blok B and Wen X-G 1991 *Phys. Rev. B* **43** 8337
- [8] Willett R L *et al* 1987 *Phys. Rev. Lett.* **59** 1776
- [9] Haldane F D M and Rezayi E H 1988 *Phys. Rev. Lett.* **60** 956, 1886(E)
- [10] Halperin B I, Lee P A and Read N 1993 *Phys. Rev. B* **47** 7312
- [11] Jain J K 1989 *Phys. Rev. Lett.* **63** 199; 1989 *Phys. Rev. B* **40** 8079; 1990 *Phys. Rev. B* **41** 7653
- [12] Rezayi E and Read N 1994 *Phys. Rev. Lett.* **72** 900
- [13] Read N 1989 *Phys. Rev. Lett.* **62** 86
- [14] Moore G and Read N 1991 *Nucl. Phys. B* **360** 362
- [15] Girvin S M and Jach T 1984 *Phys. Rev. B* **29** 5617
- [16] Wilczek F 1982 *Phys. Rev. Lett.* **49** 957
- [17] Zhang S-C, Hansson T H and Kivelson S 1989 *Phys. Rev. Lett.* **62** 82
- [18] Laughlin R B 1988 *Phys. Rev. Lett.* **60** 2677
- [19] Zhang S-C 1992 *Int. J. Mod. Phys. B* **6** 25
- [20] Greiter M and Wilczek F 1992 *Nucl. Phys. B* **374** 567
- [21] Du R-R *et al* 1993 *Phys. Rev. Lett.* **70** 2944
- [22] Willett R L *et al* 1993 *Phys. Rev. B* **47** 7344; 1993 *Phys. Rev. Lett.* **71** 3846
- [23] Kang W *et al* 1993 *Phys. Rev. Lett.* **71** 3850

Inelastic light scattering in the regimes of the integer and fractional quantum Hall effects

A Pinczuk, B S Dennis, L N Pfeiffer and K W West

AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

Abstract. Recent resonant inelastic light scattering experiments in the fractional quantum Hall regime highlight the power of the method in studies of electron–electron interactions in semiconductors of reduced dimensions. We review here light scattering experiments that determine gap excitations in the regimes of the integer and fractional quantum Hall effects. At integer values of the Landau level filling factor we consider inter-Landau level excitations. In the fractional quantum Hall regime we discuss the determination of long-wavelength gap excitations and spin waves in the incompressible quantum fluid at filling factor 1/3.

The remarkable quantum Hall effects displayed by the two-dimensional electron gas in a large perpendicular magnetic field continue to attract intense experimental and theoretical interest [1]. The transport anomalies of the integer quantum Hall effect [2], at integer values of the Landau level filling factor v , can be regarded as arising from a sequence of metal–insulator transitions in the 2D electron gas. Such evolution of the 2D electron system is explained by forbidden energy gaps that separate occupied from empty states at integer values of v . In a single-particle picture, which requires corrections due to many-electron interactions, the energy gaps arise from quantization of kinetic energy into Landau levels and the Zeeman splitting of the spin states [3, 4]. The similarity of the transport anomalies of the fractional quantum Hall effect [5], seen at ‘magic’ fractional values of v , with those measured at integer v are evidence of the emergence of unexpected energy gaps in the 2D system when free electrons reside in partially occupied Landau levels. Because of the absence of single-particle gaps at fractional v , the energy gaps of the fractional quantum Hall effect (FQHE) are a manifestation of novel fundamental behaviours of electron–electron interactions in two dimensions [4].

The energy gaps of the quantum Hall states are represented by neutral density excitations of the 2D electron gas in a large perpendicular magnetic field [6–10]. The excitations are constructed from particle–hole pair states like that shown in figure 1(a), where the particles can be either electrons or quasiparticles. The study of these ‘gap excitations’ is the principal task of spectroscopy in the quantum Hall regimes. The quest for the direct observation of gap excitations acquires prominent significance in studies of the fractional quantum Hall effect (FQHE). The emergence of new, low-lying, gap excitations in the states of the FQHE is

entirely due to electron–electron interactions. The energies and wavevector dispersion of these modes are a primary manifestation of the correlations that lead to condensation of the electron system into an incompressible quantum fluid. These gap excitations are associated with fractionally charged quasiparticles that obey fractional statistics [11, 12]. The calculated mode dispersions exhibit ‘magnetoroton’ minima that are characteristic of electron–electron interactions of 2D systems in a perpendicular magnetic field [6–10, 13].

We recently reported the direct measurement of collective excitations of the 2D electron gas in the FQHE state at $v = 1/3$ by resonant inelastic light scattering [14]. These results exhibit a sharp peak which is interpreted as a $q = 0$ gap excitation of the FQHE state. The observations are surprising because the dynamical structure factor, the function that enters in conventional inelastic scattering cross sections, vanishes for the small wavevectors in optical experiments [8, 9]. The light scattering results are also intriguing because the character of $q = 0$ gap excitations in the FQHE is still a subject of debate [8, 15]. In fact, speculation that the $q = 0$ mode could consist of a pair of excitations, each near the magnetoroton minimum, stimulated the light scattering measurements that culminated in the observations of gap excitations [14].

The results obtained in the regime of the FQHE highlight applications of the light scattering method in studies of excitations of the 2D electron gas that are not easily accessible by other spectroscopic methods [15–18]. In this paper we present a review of inelastic light scattering in the quantum Hall regimes. We briefly consider neutral density excitations and mechanisms of resonant inelastic light scattering. Next we discuss some significant results obtained at integral values of v and the more recent measurements at $v = 1/3$.

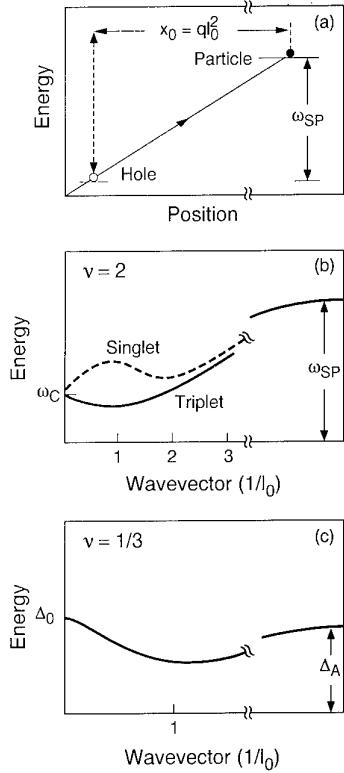


Figure 1. Neutral density excitations in the quantum Hall regimes. (a) Particle–hole pairs that take part in excitations with wavevector q . (b) Wavevector dispersion for inter-Landau-level excitations at $v = 2$. The singlet is the magnetoplasmon or charge density excitation and the triplet is a spin density excitation. (c) Calculated dispersion of charge-density intra-Landau-level excitations in the FQHE at $v = 1/3$ (after [8]). Δ_A is the energy of a quasiparticle–quasihole pair at large x_0 .

In neutral density excitations the particle is promoted to an empty state leaving behind an oppositely charged hole. In the Landau gauge, used to describe the excitations by a wavevector q , the displacement of the cyclotron orbits is $x_0 = q l_0^2$ as shown in figure 1(a). Here $l_0 = (\hbar c/eB)^{1/2}$ is the magnetic length and the direction of x_0 is orthogonal to that of q . At integer v the particle is an electron and in the FQHE the pairs consist of quasiparticles and quasiholes that carry fractional charges [6, 9]. We specialize to the case of 2D electron systems in GaAs layers, where spin is $1/2$ and orbital angular momentum is zero. The excitations can be classified according to the total angular momentum of the particle–hole pairs [6, 16]. At $v = 2$ the two spin states of the lowest Landau level are fully occupied and the pair states are either singlets ($J = 0$) or triplets ($J = 1$). In the spin-polarized case at $v \leq 1$ the excitations are either non-spin-flip or spin-flip modes. The singlets and non-spin-flip modes are charge density excitations in which there is no change in the spin degree of freedom.

Within the time-dependent Hartree–Fock approximation the excitation energies can be written as [6, 19]

$$\omega(q, B) = \omega_{\text{SP}}(B) + E_d(q, B) + E_x(q, B) \quad (1)$$

where $E_d(q, B)$ and $E_x(q, B)$ represent the couplings between particle and hole arising from the direct and exchange terms of electron–electron interactions. E_d is related to the macroscopic electric field that occurs in charge density excitations. E_x represents a binding energy of the particle–hole pair. The two terms have pronounced q dependences and vanish for $q \rightarrow \infty$, because the Fourier components of the Coulomb interaction go as $1/q$. The q -independent ω_{SP} is regarded as a ‘single-particle’ energy given by

$$\omega_{\text{SP}}(B) = \omega_0(B) + \Sigma(B) \quad (2)$$

where $\omega_0(B)$ is the transition energy and $\Sigma(B)$ is a self-energy due to changes in Coulomb interactions when an electron changes state.

We consider next inter-Landau-level excitations with Landau level index changes of $\Delta l = 1$. Figure 1(b) is a theoretical prediction for $v = 2$. The modes are gap excitations for this quantum Hall state. Here $\omega_0 = \omega_c$, the cyclotron energy of an electron with effective mass m^* . At $q = 0$ the energies of charge density excitations (or magnetoplasmons) and spin density modes are equal because $\Sigma(B)$ is the same for the two modes. The $q = 0$ spin-flip modes and magnetoplasmons are not degenerate at $v = 1$ [17, 20] (as will be seen in figure 3(b)) because in the spin-polarized electron gas the self-energies are strongly dependent on changes in the spin degree of freedom. Figures 1(b) and 3(b) display roton (or magneto-roton) minima at wavevectors $q \gtrsim 1/l_0$. The rotons, caused by the reduction in ‘excitonic’ binding $E_x(q, B)$ at wavevectors $q \gg 1/l_0$, are among the major predictions of current theories of collective excitations in the quantum Hall regime.

Figure 1(c) delineates the gap excitations of the FQHE state at $v = 1/3$ as calculated within the single-mode approximation (SMA) [8]. The modes are intra-Landau-level charge density excitations associated with quasiparticle–quasihole pairs [6–9]. The wavevector dispersion shows a pronounced magnetoroton minimum at $q \sim 1/l_0$ caused by the q dependence of an ‘excitonic’ binding energy. The $q \rightarrow \infty$ limit, which represents infinitely separated quasiparticle–quasihole pairs, is associated with the activation energy Δ_A (as indicated in figure 1(a)) [6]. Optical experiments typically access long-wavelength excitations. However, intra-Landau-level excitations have vanishing oscillator strength for $q \rightarrow 0$. Girvin *et al* [8] and, more recently, Lee and Zhang [15] conjectured that the $q = 0$ mode consists of a two-excitations state. In this picture light scattering mechanisms by the $q = 0$ gap mode could be similar to those of second-order Raman scattering.

Resonant inelastic light scattering was proposed in 1978 as a method for study of 2D electron systems in semiconductors [21]. During the last 15 years the method has become a major spectroscopic tool for studies of electrons in heterojunctions and quantum wells [18, 22–24]. While detailed evaluations of light scattering cross sections in the quantum Hall regimes are not available at present, some general behaviours can be

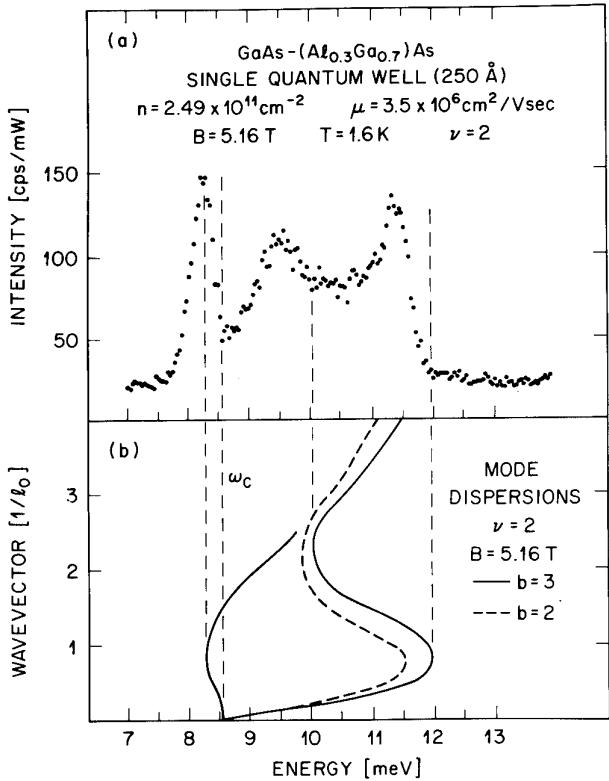


Figure 2. (a) Inelastic light scattering spectra of inter-Landau-level excitations at $\nu = 2$. (b) Calculated mode dispersions. Here b parametrizes the finite width of the electron wavefunction in the perpendicular direction (after [16]).

anticipated by examination of the scattering matrix element within the framework of the effective-mass approximation [18, 22–26]. It is found that scattering cross sections could be small in the lowest-order approximation. This is the case of inter-Landau-level excitations with change in Landau level index $\Delta l = 1$, which may have a vanishing scattering matrix element in the dipole approximation. For such excitations, mixing between Landau levels, which is significant in the valence states of GaAs quantum wells [27], results in non-vanishing matrix elements.

Typical results for inter-Landau-level excitations at $\nu = 2$ are shown in figure 2. The spectrum of figure 2(a), measured in a modulation doping single GaAs-AlGaAs quantum well, consists of three relatively broad but clearly defined intensity maxima. Figure 2(b) displays the calculated mode dispersions within the time-dependent Hartree-Fock approximation. The parameter b of the calculation is used to take into consideration the effect of the finite width of the electron wavefunctions along the direction perpendicular to the layer [17, 20]. The three structures resolved in the spectra have energies close to the positions of the critical points in the mode dispersions. These are the energies of the peaks in the density of states of the modes because $d\omega/dq = 0$. Given that the critical points occur at wavevectors $q \gtrsim 1/l_0$, much larger than the in-plane component of the light

scattering wavevector k , the results in figure 2 imply a massive breakdown of wavevector conservation. This has been associated with the loss of translational symmetry at $\nu = 2$ when the 2D electron system, being in an insulating state, is not very effective in screening the residual disorder [16].

The impact of breakdown of wavevector conservation on light scattering spectra measured at $\nu = 2$ was evaluated by Marmorkos and Das Sarma [28]. In this calculation a Lorentzian function where

$$f(k, q, \alpha) = \frac{\alpha}{(k - q)^2 + \alpha^2} \quad (3)$$

replaces the conventional wavevector conservation condition $f(k, q, 0) = \delta(k - q)$. In equation (3) the adjustable parameter α is a measure of the extent of breakdown of wavevector conservation. This approach yields calculated spectra with features at the critical points of the density of states. The relative intensity of the characteristic doublet associated with the roton is determined by the value of α . The calculations of Marmorkos and Das Sarma indicate that results like those in figure 2(a) can be explained with $\alpha = 1/d_s$, where d_s is the width of the spacer layer that separates the doping layer of ionized impurities in the AlGaAs barrier from the 2D electron gas in the GaAs quantum well. These results suggest that the residual disorder in these experiments is due to the ionized impurities of the modulation doping procedure.

The results in figure 2 exemplify the light scattering studies of gap excitations in the quantum Hall regimes. These investigations yield direct evidence of magneto-rotons in the mode dispersions, a major feature of theories of electron-electron interaction in the 2D electron gas. The two peaks seen at $\omega > \omega_c$ correspond to the multiple critical points associated with the roton minimum in the dispersion of magnetoplasmons (the singlet states in figure 1(b)). There are relatively small (0.5 meV) differences between measured and calculated positions of the critical points in the dispersion. The discrepancy could be explained in part by the strong field approximation ($\omega_c \gg e^2/\epsilon_0 l_0$) used in the calculation. Such an approximation ignores the coupling to higher inter-Landau-level transitions ($\Delta l = 2$) that at these relatively small fields could reduce the energies of $q \neq 0$ magnetoplasmons [19].

Figure 3 shows results for inter-Landau-level excitations near $\nu = 1$. Figure 3(a) presents spectra measured for several incident photo energies $\hbar\omega_L$ in resonance with excitonic transitions associated with higher Landau levels of the lowest confined state of the quantum well [16]. Figure 3(b) shows mode dispersions calculated within the time-dependent Hartree-Fock approximation [17, 20]. The calculation indicates that the strongest peaks in the spectra are the $q = 0$ magnetoplasmons (at $\omega_c = 17.7$ meV) and spin-flip inter-Landau-level excitations (at $\omega_{SF} = 22.3$ meV). The weaker structures centred at $\omega = 19.8$ meV are interpreted as roton densities of states of magnetoplasmons, observed here, as at $\nu = 2$, because

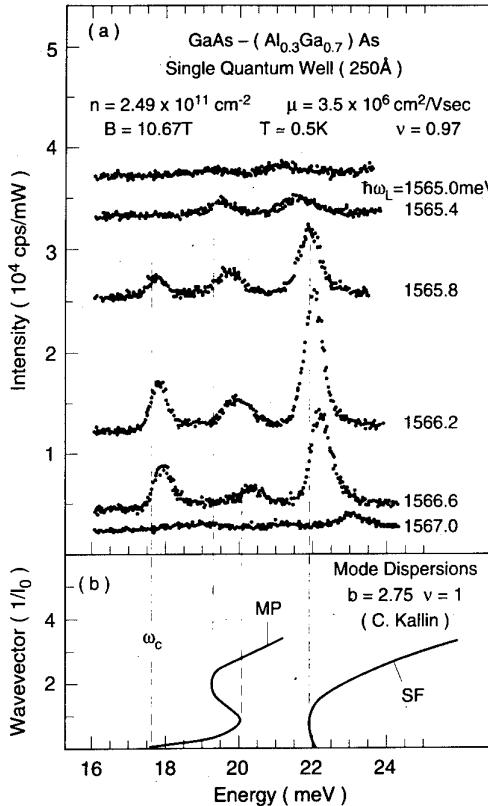


Figure 3. Light scattering spectra of inter-Landau-level excitations at four incident photon energies $\hbar\omega_L$ (after [16]). (b) Calculated mode dispersions at $v = 1$ (Kallin private communication).

of breakdown of wavevector conservation due to loss of translational invariance in the insulating state with $v = 1$.

The splitting between $q = 0$ magnetoplasmons and spin-flip inter-Landau-level excitations at $v = 1$ is related to the exchange self-energy that enters in the enhancement of the spin gap in the spin-polarized 2D electron gas [17, 20]. From equations (1) and (2) we write

$$\Delta_{SF} = E_Z + \Delta\Sigma \quad (4)$$

where E_Z is the Zeeman splitting and $\Delta\Sigma$ is the difference in exchange self-energies between spin-flip and non-spin-flip transitions. In the exact 2D limit ($b \rightarrow \infty$) [17, 20]

$$\Delta\Sigma = -\frac{1}{2}\Sigma_0 \quad (5)$$

where $-\Sigma_0$ is the self-energy that accounts for the well known enhancement of the spin gap [3].

The light scattering determinations, like that shown in figure 3, give direct spectroscopic evidence of an exchange enhancement in the spin-polarized electron gas in agreement with the prediction of the Hartree-Fock approximation. We also find that the spectra have a striking dependence on temperature that is not observed for $v = 2$. At $T = 5 \text{ K}$ the spectral intensity decreases by a factor of about 10 and the modes cannot be measured above 10 K. It is interesting that such temperature dependence is similar to that of the occupations of the two lowest spin-split Landau levels as determined by interband optical absorption [29]. This suggests that the

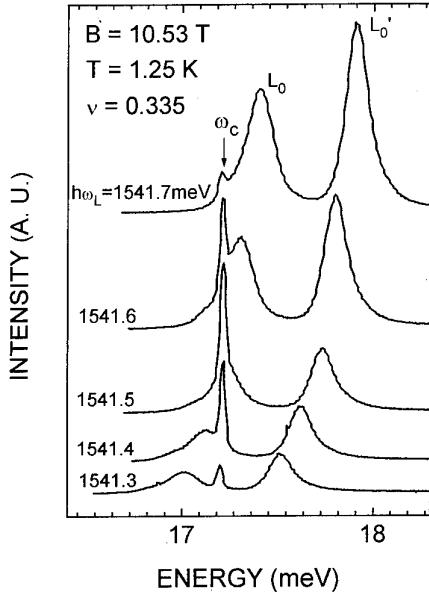


Figure 4. Light scattering by the $q = 0$ magnetoplasmon (at ω_c) for several incident photon energies. L_0 and L'_0 are the characteristic doublets of photoluminescence near the fundamental gap. GaAs/Al_{0.1}Ga_{0.9}As SQW (25 nm), $n = 8.5 \times 10^{10} \text{ cm}^{-2}$, $\mu = 3 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

temperature dependence of the light scattering intensities is related to the loss of spin-polarization in the electron gas.

The light scattering measurements of collective excitations at $v = 1/3$ have the benefit of sharp resonant enhancements of the cross sections [14]. Figure 4 shows spectra measured with scattered photon energies close to the fundamental interband energy gap of the GaAs quantum well. The bands labelled L'_0 and L_0 , with FWHM $\simeq 0.2 \text{ meV}$, are the characteristic doublets of intrinsic photoluminescence due to recombination of the electron gas with weakly photoexcited holes in the highest valence Landau level of the GaAs quantum well [30]. The sharp light scattering peak at ω_c , with FWHM $\simeq 0.02 \text{ meV}$, is interpreted as the $q = 0$ magnetoplasmon [20, 31]. Remarkably, figure 4 reveals that the peak intensity of the light scattering feature is comparable to that of intrinsic photoluminescence at the fundamental gap. These results are a striking example of the very sharp resonant enhancements that are possible in very high quality GaAs quantum wells.

Figures 5 and 6 show resonant inelastic light scattering by low-lying excitations ($\omega_c \gg \omega$) of the FQH state at $v = 1/3$ [14]. To carry out these measurements a ³He cryostat was inserted into the cold bore of a superconducting magnet. The system has silica windows for optical access. The conventional backscattering geometry allows only small values $k \lesssim 10^4 \text{ cm}^{-1}$ of the light scattering wavevector [22-24]. The experimental configuration, in conjunction with almost perfect sample surfaces, is associated with the low stray light that enables measurements at energy shifts as small as 0.2 meV. Here the power densities of the incident laser light are about

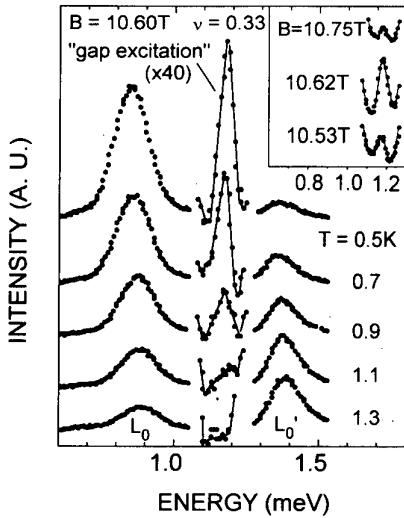


Figure 5. Temperature dependence of light scattering spectra of a low-lying excitation of the FQHE at $v = 1/3$. The inset shows the B dependence of the 0.5 K spectra. The temperature dependence of the L_0 and L'_0 intensities is due to the optical anomaly at $v = 1/3$ (after [14]). GaAs/Al_{0.1}Ga_{0.9}As SQW (25 nm), $n = 8.5 \times 10^{10} \text{ cm}^{-2}$, $\mu = 3 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

$10^{-4} \text{ W cm}^{-2}$, and the photon energies are resonant with lowest optical transitions of the quantum well. The spectra were recorded with optical multichannel detection at a resolution of $\sim 0.02 \text{ meV}$. The sharp peaks at 1.18 meV in figure 5 and at 0.25 meV in figure 6 are due to $q \sim 0$ excitations. This follows from the sharpness of the peaks, which indicates that wavevector is conserved for light scattering by these excitations ($k = q \sim 0$). If this were not the case much larger widths would be observed, as in the cases illustrated in figures 2 and 3.

The light scattering intensities in the spectra of figures 5 and 6 have a striking temperature dependence. In the case of the peak at 1.18 meV we also find a marked dependence on magnetic field, in which the mode is observed only in the small interval $\Delta B \simeq 0.5 \text{ T}$ near $v = 1/3$. Such field dependence is shown in the inset to figure 5. These temperature and magnetic field dependences are characteristic of the fractional quantum Hall effect [5, 32]. They indicate that the sharp peaks are due to long-wavelength intra-Landau-level excitations with $\Delta l = 0$ in the FQHE state of $v = 1/3$.

The peak at 1.18 meV has been interpreted as a $q \sim 0$ gap excitation of the incompressible state at $v = 1/3$. For a rigorous comparison with theoretical predictions we need to incorporate the effect of the finite width of the 2D layer. Its impact on the $q = 0$ gap excitation has been evaluated within a single-mode approximation [8]. Interpolation in the calculation of Girvin *et al* [8] indicates good agreement between the experimental result and the theoretical prediction. The peak at 0.25 meV has been assigned to the $q = 0$ collective spin wave in the lowest Landau level because its energy can be understood as a Zeeman splitting with $g = 0.43$. This value of g -factor is close to those reported for the 2D electron gas in GaAs heterojunctions [33].

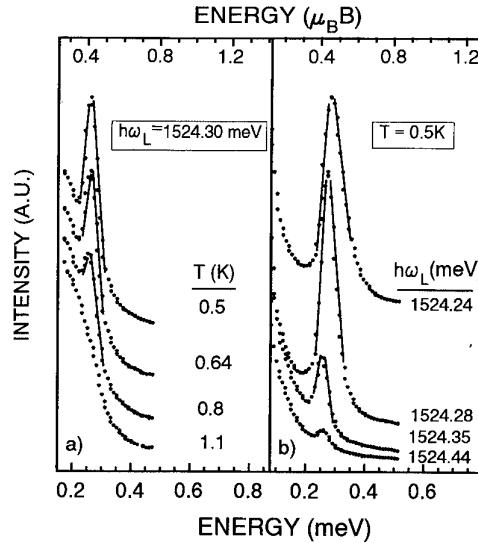


Figure 6. Light scattering spectra of a low-lying excitation at $v = 1/3$. The sharp peak is identified as a $q = 0$ spin-wave excitation. The background is due to L_0 luminescence. (a) Temperature dependence; (b) dependence on $\hbar\omega_L$ (after [14]). GaAs/Al_{0.1}Ga_{0.9}As SQW (25 nm), $n = 8.5 \times 10^{10} \text{ cm}^{-2}$, $\mu = 3 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, $B = 10.60 \text{ T}$, $v = 0.33$.

The marked temperature dependence of the $q \sim 0$ gap excitation is consistent with magnetotransport measurements that show well defined activated behaviour only for $T \leq 1 \text{ K}$ [32, 34]. In contrast, the strong temperature dependence of light scattering by $q = 0$ spin-wave excitations is unexpected. The temperature dependence of these excitations should be governed by the Zeeman energy $E_Z = 3 \text{ K}$. We find instead that the intensities have a strong temperature dependence for $T < 1 \text{ K}$ similar to that of the FQHE.

In the more recent experiments we measured the long-wavelength gap excitation in several samples in the density range $0.7 < n < 1.4 \times 10^{11} \text{ cm}^{-2}$. While these results can be regarded as definitive evidence of light scattering observations of $q = 0$ gap modes of the FQHE state at $v = 1/3$, the light scattering mechanisms active in these experiments are not yet understood. Inelastic light scattering in the fractional quantum Hall regime should be driven, as in more conventional cases [18, 21–23], by two-photon processes that resonantly create virtual electron–hole pair states. At a given photon energy near the semiconductor bandgap, these pair states dominate the scattering matrix element because of vanishingly small resonant denominators. Light scattering by gap excitations of the FQHE is caused by interactions of these pairs with quasiparticles and quasiholes that carry fractional charges. We envision that further experimental research of light scattering resonant enhancement profiles in conjunction with theoretical models of cross sections could reveal new physics on the structure of the $q \sim 0$ gap excitations. Such studies are of considerable interest given that, as mentioned earlier, the character of the $q \sim 0$ gap excitations is still a subject of debate.

Light scattering experiments in the regime of the FQHE could also access large- q excitations through breakdown of the condition of wavevector conservation. This will occur when the emergence of an unscreened residual disorder potential in the incompressible, i.e. insulating, states of the FQHE enhances loss of translational invariance and activates light scattering by large-wavevector excitations in this scenario. As in the case of integer v , in this scenario the spectra would show structure at the energies of the maxima in the densities of states of collective excitations. In the case of gap excitations of the FQHE, such experiments could test the prediction of magnetoroton minima and other significant features in the dispersions of the modes.

Light scattering studies in the fractional quantum Hall regime are not limited to gap excitations at $v = 1/3$. Further studies should test current theoretical predictions for the dispersion of spin waves and inter-Landau-level excitations [20, 31, 35, 36]. The impact of the light scattering method could extend to other states like that at $v = 2/5$, where measurements of gap excitations could explore the new hierarchical constructions of the FQHE [34, 37, 38]. Resonant inelastic light scattering could also be used to investigate the collective excitations that might exist in the states of transformed fermions interacting with a Chern–Simons gauge field, recently proposed for the state at $v = 1/2$ [39]. Further experiments might also give insights into the electromagnetic responses of the states of the FQHE [40, 41].

Acknowledgments

We thank L Brey, E Burstein, J P Eisenstein, B I Halperin, Song He, D Heiman, C Kallin, B D McCombe, A H MacDonald, P M Platzman, L Sohn and H L Stormer for discussions on several aspects of this work.

References

- [1] Proc. 10th Int. Conf. on Electronic Properties of 2-Dimensional Systems (EP2DS-10) (1994 Surf. Sci. 305)
- [2] von Klitzing K, Dorda G and Pepper M 1980 Phys. Rev. Lett. **45** 494
- [3] Ando T, Fowler A B and Stern F 1982 Rev. Mod. Phys. **54** 457
- [4] Prange R E and Girvin S M (ed) 1990 2nd edn (New York: Springer)
- [5] Tsui D C, Stormer H L and Gossard A C 1982 Phys. Rev. Lett. **48** 1559
- [6] Kallin C and Halperin B I 1984 Phys. Rev. B **30** 5655
- [7] Haldane F D M and Rezayi E H 1985 Phys. Rev. Lett. **54** 237
Rezayi E H and Haldane F D M 1985 Phys. Rev. B **32** 6924
Haldane F D M 1990 *The Quantum Hall Effect* 2nd edn ed R E Prange and S M Girvin (New York: Springer) p 303
- [8] Girvin S M, MacDonald A H and Platzman P M 1985 Phys. Rev. Lett. **54** 581; 1986 Phys. Rev. B **33** 2481
Girvin S M 1990 *The Quantum Hall Effect* 2nd edn ed R E Prange and S M Girvin (New York: Springer) p 353
- [9] Laughlin R B 1990 *The Quantum Hall Effect* 2nd edn ed R E Prange and S M Girvin (New York: Springer) p 233
- [10] Chakraborty T and Pietilainen P 1988 *The Fractional Quantum Hall Effect* (New York: Springer)
- [11] Halperin B I 1984 Phys. Rev. Lett. **52** 1583
- [12] Arovas D, Schrieffer J R and Wilczek F 1984 Phys. Rev. Lett. **53** 722
- [13] Lerner I V and Lozovik Yu E 1980 Sov. Phys.–JETP **51** 5881
- [14] Pinczuk A, Dennis B S, Pfeiffer L N and West K 1993 Phys. Rev. Lett. **70** 3983
- [15] Lee D-H and Zhang S-C 1991 Phys. Rev. Lett. **66** 1220
- [16] Pinczuk A, Heiman D, Schmitt-Rink S, Chuang S L, Kallin C, Valladares J P, Dennis B S, Pfeiffer L N and West K W 1990 Proc. 20th Int. Conf. on the Physics of Semiconductors ed E M Anastassakis and J D Joannopoulos (Singapore: World Scientific) p 1045
- [17] Pinczuk A, Dennis B S, Heiman D, Kallin C, Brey L, Tejedor C, Schmitt-Rink S, Pfeiffer L N and West K W 1992 Phys. Rev. Lett. **68** 3623
- [18] Pinczuk A, Dennis B S, Pfeiffer L N, West K W and Burstein E 1994 Phil Mag. B at press
- [19] MacDonald A H 1985 J Phys. C: Solid State Phys. **18** 103
- [20] Longo J P and Kallin C 1993 Phys. Rev. B **47** 4429
- [21] Burstein E, Pinczuk A and Buchner S 1979 *Physics of Semiconductors* 1978 B L H Wilson (London: Institute of Physics) p 1231
- [22] Abstreiter G, Cardona M and Pinczuk A 1984 *Light Scattering in Solids IV* ed M Cardona and G Guentherodt (Berlin: Springer) p 5
- [23] Abstreiter G, Merlin R and Pinczuk A 1986 IEEE J. Quantum Electron. **22** 1771
- [24] Pinczuk A 1992 *Festkörperprobleme/Advances in Solid State Physics* vol 32 ed U Roessler (Braunschweig: Vieweg) p 45
- [25] Wolff P A 1966 Phys. Rev. Lett. **16** 225
- [26] Hamilton D C and McWhorter A L 1969 *Light Scattering Spectra of Solids* ed G B Wright (New York: Springer) p 309
- [27] Goldberg B B, Heiman D, Graf M J, Broido D A, Pinczuk A, Tu C W, English J H and Gossard A C 1988 Phys. Rev. B **38** 10131
- [28] Marmorkos I K and Das Sarma S 1992 Phys. Rev. B **45** 13396
- [29] Goldberg B B, Heiman D and Pinczuk A 1989 Phys. Rev. Lett. **63** 1102
- [30] Heiman D, Goldberg B B, Pinczuk A, Tu C W, Gossard A C and English J H 1988 Phys. Rev. Lett. **61** 605
- [31] MacDonald A H, Oji H C A and Girvin S 1985 Phys. Rev. Lett. **55** 2208
- [32] Boebinger G S, Chang A M, Stormer H L and Tsui D C 1985 Phys. Rev. Lett. **55** 1606
Willett R L, Stormer H L, Tsui D C, Gossard A C and English J H 1988 Phys. Rev. B **37** 8476
- [33] Dobers M, von Klitzing K and Weimann G 1988 Phys. Rev. B **38** 5453
- [34] Du R R, Stormer H L, Tsui D C, Pfeiffer L N and West K W 1993 Phys. Rev. Lett. **70** 2944
- [35] Rasolt M and MacDonald A H 1984 Phys. Rev. B **30** 5655
- [36] Pietilainen P and Chakraborty T 1988 *Europhys. Lett.* **5** 157
- [37] Jain J K 1989 Phys. Rev. Lett. **63** 199
- [38] Kang W, Stormer H L, Pfeiffer L N, Baldwin K W and West K W 1993 Phys. Rev. Lett. **71** 3850
- [39] Halperin B I, Read N and Lee P A 1993 Phys. Rev. B **47** 7312
- [40] Lopez A and Fradkin E 1993 Phys. Rev. B **47** 7080
- [41] Simon S H and Halperin B I 1993 Phys. Rev. B **48** 17386

Interactions and transport in nanostructures

B Kramer, T Brandes, W Häusler, K Jauregui, W Pfaff
and D Weinmann

I. Institut für Theoretische Physik, Universität Hamburg, Jungiusstrasse 9,
D-20355 Hamburg, Germany

Abstract. We discuss various aspects of electron–electron and electron–phonon interaction in electron transport in submicrometre structures. We show that it is only above a certain critical Fermi velocity that the acoustic phonons can significantly influence the electron states in a quasi-one-dimensional quantum wire. We predict a characteristic temperature dependence of the plateaus in the linear conductance as a function of a magnetic field which should be experimentally observable.

When the mean distance between Coulombically interacting electrons in a quantum dot is comparable to or larger than the Bohr radius their excitation spectrum shows fine structure which is related to the formation of a localized charge distribution, a Wigner molecule. We demonstrate that the excitations can be understood in terms of vibrational and tunnelling modes. Nonlinear transport of confined interacting electrons coupled to semi-infinite leads yields detailed information about the excitation spectrum. We present results including the degrees of freedom that were obtained from a master equation approach, and demonstrate that the correlations between the electrons lead to negative differential resistances that are related to spin selection rules.

1. Introduction

Interaction processes are of great importance for the understanding of the electronic transport properties of quantum coherent samples, although the conditions under which such a system can be considered as coherent depend strongly on the *absence* of phase-breaking scattering. Among the many possibilities, electron–phonon (e–p) and electron–electron (e–e) interaction are most important [1]. In this contribution we discuss three aspects which demonstrate some most striking features in connection with electron transport in submicrometre structures.

First, we consider the interaction of acoustic phonons with the electrons in a quasi-one-dimensional quantum wire. We will demonstrate that it is only above a certain critical Fermi velocity that the phonons can influence the electron states significantly [2]. As a consequence, we predict a characteristic temperature dependence of the plateaus in the linear conductance as a function of a magnetic field which should be experimentally observable.

The Coulomb interaction between the electrons in a system that contains a quantum dot leads to rather sharp and regular resonance-like peaks in the linear conductance [3]. When the mean distance between the electrons in the dot is comparable to the Bohr radius (of the order of 10 nm for AlGaAs based heterostructures) the excitation spectrum of the interacting electrons shows fine structure which is related to the formation of a localized charge distribution, a Wigner molecule [4]. We show that the

excitation spectrum can be understood in terms of vibrational and tunnelling modes [5].

Nonlinear transport properties of interacting electrons in a quantum dot that is coupled to semi-infinite leads yield detailed information about the excitation spectrum. Using a master equation we present an investigation which includes also the spin degrees of freedom of the electrons [6, 7]. We demonstrate that correlations between the electrons lead to novel effects in the current. Negative differential resistances related to spin selection rules are predicted in accordance with recent experiments [8, 9]. Furthermore, asymmetric conductance peaks occur due to asymmetries in the coupling.

2. Acoustic phonon scattering in quantum wires

In very pure samples, e–e and e–p interaction remain the only scattering mechanisms of importance at low temperatures. We consider here scattering with acoustical phonons with speed of sound c_s . In quasi-one-dimensional (1D) ‘metallic’ electron systems based on semiconductor heterostructures the e–p interaction is qualitatively different from that in metals since in a quasi-1D subband the Fermi velocity can be made arbitrarily small, even smaller than the sound velocity, by carefully adjusting a magnetic field or a gate voltage.

We show that momentum and energy conservation lead to a crossover between two different e–p scattering regimes when the Fermi velocity v_F is tuned from small

to higher values and eventually becomes larger than c_s . A similar mechanism has been discussed some time ago in the context of the breakdown of the quantum Hall effect [10, 11]. In a quantum wire, the linear conductance Γ is determined by the scattering of incoming electrons [12] having Fermi velocity v_F . In ballistic transport the velocity does not affect the conductance, owing to a cancellation of v_F with the 1D density of states (DOS) up to a factor $2\pi\hbar$, leading to a quantization of Γ in units of e^2/h . When scattering occurs the value of v_F plays a crucial role for scattering rates, and therefore for the conductance itself.

The quantum wire is modelled in the x - y plane by using a parabolic confinement potential $V(y) = (1/2)m^*\omega_0^2y^2$. A magnetic field B is assumed in the z direction. In the Landau gauge the eigenstates of the unperturbed system are

$$\Psi_{nk}(x) = \frac{1}{\sqrt{L}} \exp(ikx)\chi(z)\phi_{nk}(y) \quad (1)$$

where ϕ_{nk} is the n th harmonic oscillator wavefunction displaced by $y_0 = -\hbar\omega_c/(m^*\omega_B^2)$ with effective magnetic length $l_B = (\hbar/m^*\omega_B)^{1/2}$. Here, $\omega_c = (eB/m^*c)$ is the cyclotron frequency (m^* effective electron mass) and $\omega_B = (\omega_0^2 + \omega_c^2)^{1/2}$. The corresponding energy dispersion is $\epsilon_{nk} = \epsilon_n + \gamma_B(\hbar^2/2m^*)k^2$ with $\gamma_B = (\omega_0/\omega_B)^2$ and $\epsilon_n = (n + 1/2)\hbar\omega_B$. The factor γ_B can become much less than unity for strong magnetic fields.

We consider longitudinal acoustic phonons with frequency ω_Q and linear dispersion relation $\omega_Q = c_s|\mathbf{Q}|$, \mathbf{Q} being the 3D phonon wavevector. Optical phonons are not included. Their energy is too large to be important at the low temperatures considered here. We restrict ourselves to the lowest electronic subband $n = 0$.

The interaction Hamiltonian is

$$H_{e-p} = \frac{1}{\sqrt{\Omega}} \sum_{k,k',Q} V_Q M_{k,k'}(\mathbf{Q}) c_k^\dagger c_{k'}(a_Q + a_{-Q}^\dagger) \quad (2)$$

with matrix elements $M_{k,k'}(\mathbf{Q}) = \langle k | \exp(i\mathbf{x}\mathbf{Q}) | k' \rangle$, and electron–phonon matrix elements $V_Q = D(\hbar Q/2\rho_M c_s)^{1/2}$. D denotes the deformation potential constant, ρ_M is the mass density and Ω the system volume.

Momentum conservation for the scattering of an electron from $|k\rangle$ to $|k'\rangle$ requires $\mathbf{Q}_x = k - k'$ and therefore $|\mathbf{Q}| \geq |\mathbf{Q}_x| = |k - k'|$. Energy conservation yields $\gamma_B(\hbar/2m^*)|k^2 - k'^2| = c_s|\mathbf{Q}| \geq c_s|k - k'|$, from which $\gamma_B(\hbar/2m^*)|k + k'| \geq c_s$ follows. For small momentum transfers we have $k \approx k'$, and we can evaluate the last inequality assuming $k \approx k_F$. It follows that $v_F \geq c_s$ is the condition for such a scattering event to take place.

We predict that the crossover mentioned above should be in principle observable in the two-point conductance of a quasi-ballistic quantum wire below about 1 K provided that e–p scattering is sufficiently significant at low temperatures. By tuning the Fermi velocity one can go from a region of exponentially small e–p scattering rates to a region with rates $\sim T^3$ which should lead to a corresponding decrease of the conductance plateau if $v_F > c_s$.

We assume the e–p interaction to be sufficiently weak to be treated by perturbation theory. In GaAs/AlGaAs this is certainly a good approximation. The electronic self-energy can be calculated by standard techniques to second order in V_Q . The imaginary part gives the inelastic lifetime of an electron of momentum k and energy ϵ

$$\begin{aligned} \tau^{-1}(\epsilon, k) = & \frac{2\pi}{\hbar\Omega} \sum_{Q,k} |V_Q|^2 |M_{k,k'}(\mathbf{Q})|^2 \\ & \times [(n_Q + f(\epsilon_{k'}))\delta(\epsilon - \epsilon_{k'} + \hbar\omega_Q) \\ & - (\omega_Q \rightarrow -\omega_Q)] \end{aligned} \quad (3)$$

Here, n_Q denotes the Bose function for phonon frequency ω_Q and f the Fermi function with chemical potential μ . The terms in the square brackets correspond to phonon emission ($+\hbar\omega_Q$) and absorption ($-\hbar\omega_Q$), respectively.

Although electronic transport properties are described by the transport rate that has in general a different temperature dependence, the scattering rate τ^{-1} should reveal at least qualitatively the influence of e–p scattering [1]. If we found an exponentially small scattering rate we would conclude that the change of the conductance is also negligibly small. On the other hand, in the region where $\tau^{-1} \sim T^3$ we would expect a large influence on the conductance, though the quantitative T dependence might not be correct.

The matrix element $M_{k,k'}$ implies momentum conservation in the longitudinal (x) direction

$$|M_{k,k'}(\mathbf{Q})|^2 = \delta_{k,k'+Q_x} |\langle \chi \phi_{0k} | \exp(i\mathbf{x}_\perp \mathbf{Q}_\perp) | \chi \phi_{0k'} \rangle|^2 \quad (4)$$

and reflects the fact that in the transverse direction the system is not translationally invariant.

Using the above definitions the scattering rate can be calculated. It consists of two contributions

$$\tau^{-1}(\epsilon, k) = \tau_+^{-1}(\epsilon, k) + \tau_-^{-1}(\epsilon, k) \quad (5)$$

where \pm correspond to scattering to positive and negative values of the final momentum k_ω .

It is lengthy but straightforward to show that at low temperatures $k_B T \ll \epsilon_F - \epsilon_0$, and for $v_F > c_s$, $\tau_+^{-1}(\epsilon_F) \sim T^3$ as for e–p scattering in metals. For $v_F < c_s$ only phonon absorption is possible and

$$\tau_+^{-1}(\epsilon_F) \sim T \exp[-2(1 - v_F/c_s)\beta m^* c_s^2].$$

The rate τ_-^{-1} in this region is exponentially small, independent of the velocity ratio, so that the total scattering rate is essentially due to scatterings that do not cross the sample. Their efficiency in reducing the conductance Γ is less than that of backscattering across the wire. They nevertheless provide an important mechanism for breaking the phase coherence of the electrons which also leads to a reduction of Γ . In addition, in a strong magnetic field scattering across the wire is strongly suppressed by the exponential smallness of the corresponding matrix elements $M_{k,-k_\omega}$.

The crossover starts to get smeared out if $k_B T \approx \epsilon_F - \epsilon_0$. When the Fermi velocity is tuned by a gate voltage the temperature below which the effect is predicted to be observable at $v_F = c_s$ is given by $k_B T \approx (1/2)m^*c_s^2$. An estimate for GaAs with $m^* = 0.067m_e$ and $c_s = 5370 \text{ m s}^{-1}$

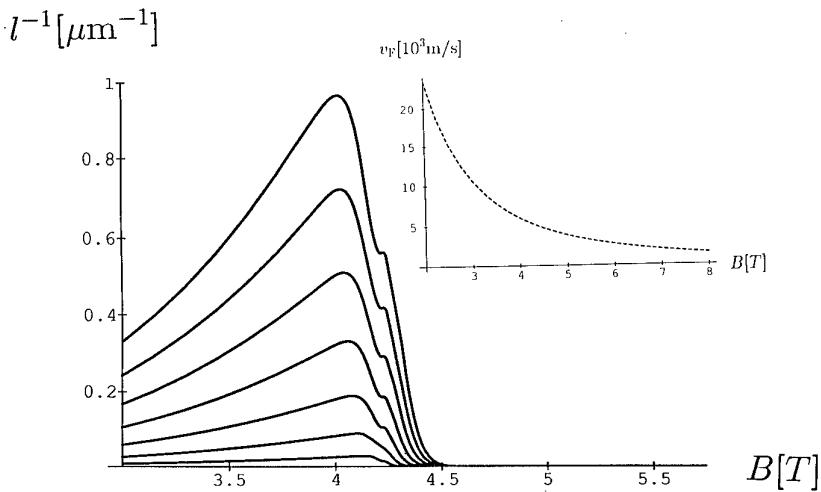


Figure 1. Inverse inelastic scattering length l^{-1} due to intraband e-p scattering in the lowest subband of a quantum wire in a magnetic field B . Temperatures are from 0.4 K to 1.6 K in steps of 0.2 K (from below), parameters were chosen with respect to a recent experiment [13] as $\hbar\omega_0 = 0.46$ meV and $n_L = 5 \times 10^8 \text{ m}^{-1}$. Inset: Fermi velocity as a function of B .

yields $T = 64$ mK which is very low and might prevent the effect being observed here.

The situation is, however, drastically different if the Fermi velocity is tuned by a magnetic field [13]. For the parabolic confining potential one has $\epsilon_F = \epsilon_{k_F}$ with $k_F = (\pi/2)n_L$ and $v_F = \gamma_B(\hbar k_F/m^*)$, where n_L is the linear density of the electrons. Since $\epsilon_F - \epsilon_0 = (1/2\gamma_B)m^*v_F^2$ with $\gamma_B \ll 1$ depending on the strength of the confining potential and the B field, relatively large $\epsilon_F - \epsilon_0$ can yield quite low Fermi velocities. One can easily reach $v_F = c_s$ for $B \neq 0$, although the Fermi energy (which depends on the magnetic field) is relatively large.

Figure 1 shows the inverse scattering length $l^{-1} = [v_F\tau_+(\epsilon_F)]^{-1}$ at different temperatures as a function of B for the lowest subband. The deformation potential D was assumed to be 13.5 eV, the mass density $\rho_M = 5.3 \text{ g cm}^{-3}$ and the speed of sound $c_s = 5370 \text{ m s}^{-1}$. The wavefunction perpendicular to the 2DEG plane was assumed to be a Gaussian with an effective width of 100 Å. The results are not too sensitive to the precise value of this parameter. One clearly observes the onset of the scattering for $v_F > c_s$. At lower B , the inverse scattering length decreases because the DOS becomes smaller†.

3. Electron interaction in quantum dots

The importance of e-e scattering for transport in quantum wires is not yet very well established experimentally, although there are strong theoretical indications that the interplay of e-e exchange interaction and confinement energy in strong magnetic fields determines whether or not the conductance is quantized [14] in terms of e^2/h or $2e^2/h$. On the other hand, there are strong

indications that Coulomb interactions are crucial for a thorough understanding of transport through quantum dots. For instance, periodic oscillations of the conductance through quantum dots that are weakly coupled to leads are well established consequences of the charging energy of single electrons entering or leaving the dot at sufficiently low temperatures [3].

To be specific we considered $N \leq 4$ interacting electrons in a quasi-1D square well of the length L including the spin degree of freedom [15]. We calculated numerically the exact eigenvalues E_v and the corresponding N -electron states $|v\rangle$ using the basis of Slater determinants. The Hilbert space was restricted to the M energetically lowest one-electron states. The interaction potential $\propto [(x - x')^2 + \lambda^2]^{-1/2}$ was used, where $\lambda (\ll L)$ is a parameter which simulates a transverse spread of the N -electron wavefunction. The Hamiltonian is

$$H = E_H \frac{a_B}{L} \left(\frac{a_B}{L} H_0 + H_I \right) \quad (6)$$

with the kinetic part

$$H_0 = \sum_{n,\sigma} \epsilon_n c_{n,\sigma}^\dagger c_{n,\sigma} \quad (7)$$

($\epsilon_n \approx n^2$, n integer), and the interaction energy

$$H_I = \sum_{m_1 \dots m_4, \sigma_1, \sigma_2} V_{m_4 m_3 m_2 m_1} c_{m_4 \sigma_1}^\dagger c_{m_3 \sigma_2}^\dagger c_{m_2 \sigma_2} c_{m_1 \sigma_1}. \quad (8)$$

$E_H = e^2/a_B$ is the Hartree energy, $a_B = \epsilon \hbar^2/m^*e^2$ the Bohr radius and ϵ the relative dielectric constant. The interaction matrix elements $V_{m_4 m_3 m_2 m_1}$ are real and do not depend on the spin state. Since the interaction is spin independent, the N -electron total spin S is a good quantum number. Figure 2 shows typical spectra for various N and $L = 9.45a_B$.

For electron densities that are not too large a tendency towards Wigner crystallization is found (figure 3) [4]. In this regime, the excitation spectrum consists of

† Directly at the crossover the curves show a small shoulder. This is due to the fact that changing from phonon absorption to phonon emission, one passes through a region of zero-phonon DOS.

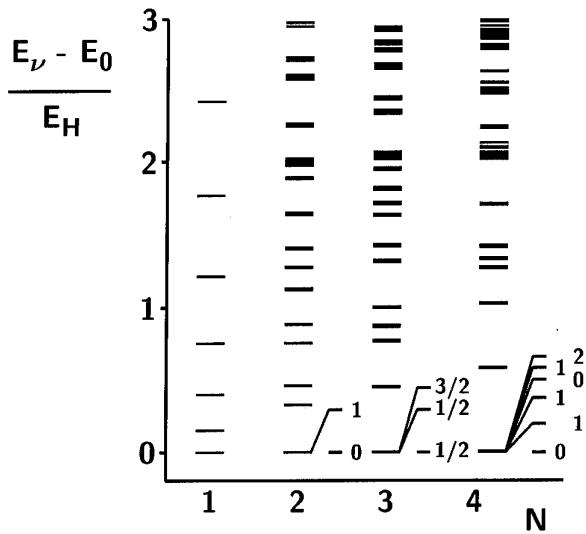


Figure 2. Energy spectra for different electron numbers for a system of length $9.45a_B$ in atomic energy units, E_H . Ground state energies E_0 are subtracted.

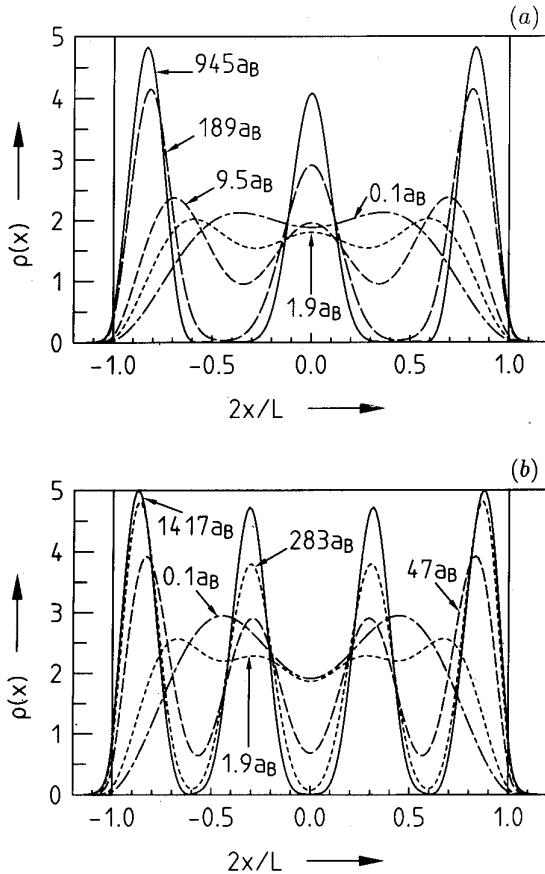


Figure 3. Charge density $\rho(x)$ for $N = 3$ (a) and $N = 4$ (b) for different L ($0.1a_B \leq L \leq 1417a_B$, $M = 13$). The normalization is such that $\int dx \rho(x) = N$. When $L \gtrsim 1a_B$ N peaks begin to emerge. For $L \gtrsim 100a_B$ the peaks are well separated.

well separated multiplets, each containing 2^N almost degenerate states [5]. The energetic differences between adjacent multiplets decrease according to a power law with electron density (figure 4). They correspond to

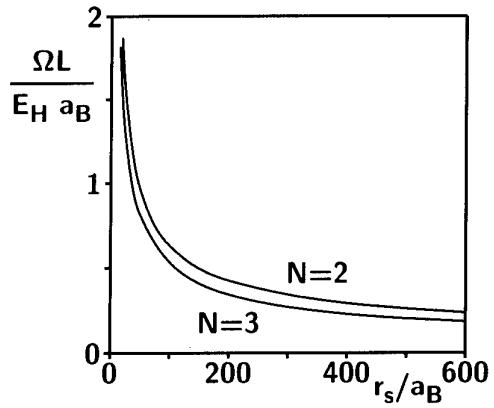


Figure 4. Energy difference between the two lowest multiplets, Ω , multiplied by L/a_B versus the mean particle distance r_s . Ω decreases more rapidly than $\sim r_s^{-1}$.

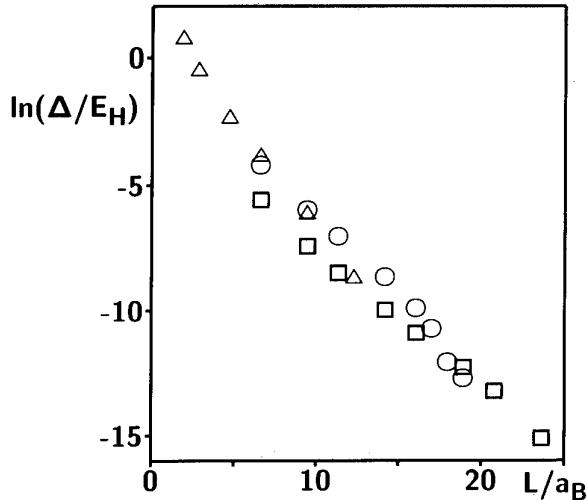


Figure 5. Logarithm of the energy difference Δ between the ground state and the first excited state within the lowest multiplet versus the system length for $N = 2$, $M = 11$ (\square), $N = 3$, $M = 13$ (\circ) and $N = 4$, $M = 10$ (\triangle). The slope defines a critical mean electron distance of $1.5a_B$, which separates the non-interacting electron spectrum from the spectrum that is characteristic for the 1D interacting electron system.

vibrational excitations. At elevated electron densities the multiplets start to split exponentially (figure 5) revealing internal fine structure. Both types of excitation energies vary with diameter L of the sample, different from the L^{-2} behaviour of non-interacting electrons. The wavefunctions of individual levels within a given multiplet differ in symmetry and S . The internal structure of the level multiplets, which form the low-energy excitations of the correlated electron system, can be understood in terms of tunnelling between various arrangements of the separated electrons [16] (cf figure 3). A formulation in terms of localized, correlated many-electron wavefunctions, in contrast to the molecular field approximation, allows us to determine the fine structure spectrum analytically for $N \leq 4$ (table 1) and enormously simplifies calculations for larger N . Generalization to 2D situations

Table 1. Spin and energies of low-lying excitations of the correlated electron model at sufficiently large electron distances $r_s \equiv L/(n-1) \gg a_B$. The tunnelling integrals t_n decrease exponentially with r_s .

n	S	$E_v - E_0(N)$
2	0	0
2	1	$2t_2$
3	1/2	0
3	1/2	$2t_3$
3	3/2	$3t_3$
4	0	0
4	1	$(1 - \sqrt{2} + \sqrt{3})t_4$
4	1	$(1 + \sqrt{3})t_4$
4	0	$(2\sqrt{3})t_4$
4	1	$(1 + \sqrt{2} + \sqrt{3})t_4$
4	2	$(3 + \sqrt{3})t_4$

is possible and the agreement with numerical calculations is convincing. The different S of excited states should in principle be detectable by ESR and play a crucial role for nonlinear transport (section 4); the state with maximal spin $S = N/2$ is never the ground state. It can be shown that the ratios between the fine structure excitation energies are ‘universally’ independent of (not too high) electron densities and of the detailed form of the e^-e^- repulsion.

At bias voltages larger than the differences between discrete excitation energies within the dot, a characteristic splitting of the conductance peaks is observed [8, 9]. We will demonstrate unambiguously below that this is related to transport involving the excited states of N correlated electrons and that the shape of the peaks depends on the coupling between the quantum dot and the leads.

4. Coulomb and spin blockade effects

We consider the double-barrier Hamiltonian

$$H = H_l + H_r + H_D + H_l^T + H_r^T \quad (9)$$

where $H_{l/r} = \sum_{k,\sigma} \epsilon_k^{l/r} c_{l/r,k,\sigma}^\dagger c_{l/r,k,\sigma}$ describes free electrons in the left/right lead and

$$H_D = \sum_{m,\sigma} (\epsilon_m - eV_G) c_{m,\sigma}^\dagger c_{m,\sigma} + \sum_{m_1, m_2, m_3, m_4, \sigma_1, \sigma_2} V_{m_1 m_2 m_3 m_4} c_{m_1, \sigma_1}^\dagger c_{m_2, \sigma_2}^\dagger c_{m_3, \sigma_2} c_{m_4, \sigma_1} \quad (10)$$

the interacting electrons within the dot. The voltage V_G is the potential change in the dot due to an externally applied voltage and serves to change the electron density in the well.

The barriers are represented by the tunnelling Hamiltonians

$$H_{l/r}^T = \sum_{k,m,\sigma} (T_{k,m,\sigma}^{l/r} c_{l/r,k,\sigma}^\dagger c_{m,\sigma} + \text{HC}) \quad (11)$$

where $T_{k,m,\sigma}^{l/r}$ are the transmission probability amplitudes which we assume to be independent of m and σ . We assume that the phase coherence between the eigenstates of H is destroyed within a time τ_Φ , on the average, which is much larger than the time an electron needs to travel from one barrier to the other. Thus, the motion of the electrons inside the dot is sufficiently coherent to guarantee the existence of quasi-discrete levels†. We assume also that the leads are in thermal equilibrium described by the Fermi–Dirac distributions $f_{l/r}(e) = \{\exp[\beta(e - \mu_{l/r})] + 1\}^{-1}$. The chemical potential in the left/right lead is $\mu_{l/r}$ and $\beta = 1/k_B T$ the inverse temperature. We assume the tunnelling rates through the barriers $t^{l/r} = (2\pi/\hbar) \sum_k |T_{k,m,\sigma}^{l/r}|^2 \delta(\epsilon_k^{l/r} - E)$ to be independent of energy E . If they are small compared with the phase-breaking rate τ_Φ^{-1} , the time evolution of the occupation probabilities of the many-electron states in the dot can be calculated using a master equation [18–20]. We take into account the populations P_i of all possible Fock states $|i\rangle$ of H_D . Transitions between the latter occur when an electron tunnels through a barrier. Our method, which is based on the exact many-electron states of the dot including spin, allows us to determine the stationary non-equilibrium state without being restricted to the conventional charging model.

Each of the states $|i\rangle$ is associated with a certain electron number n_i , an energy eigenvalue E_i and the total spin S_i . For sufficiently small H^T , the transition rates between $|i\rangle$ and $|j\rangle$ with $N_i = N_j + 1$ are $\Gamma_{j,i}^{l/r,-}$ and $\Gamma_{i,j}^{l/r,+}$, depending on whether an electron is leaving or entering through the left/right barrier respectively. Here, $\Gamma_{j,i}^{l/r,-} = \gamma_{j,i} t^{l/r} (1 - f_{l/r}(E))$, $\Gamma_{i,j}^{l/r,+} = \gamma_{i,j} t^{l/r} f_{l/r}(E)$ and the electron provides the energy $E = E_i - E_j$. As an additional, and very important, selection rule we take into account that each added or removed electron can change the total spin S_i of the N_i electrons in the dot only by $\pm 1/2$. The consideration of the vector-coupling Clebsch–Gordan coefficients yields the spin-dependent factors

$$\gamma_{j,i} = \frac{S_i + 1}{2S_i + 1} \delta_{S_i+1/2, S_j} + \frac{S_i}{2S_i + 1} \delta_{S_i-1/2, S_j} \quad (12)$$

in the transition rates.

The master equation for the time evolution of the occupation probabilities P_i is

$$\frac{d}{dt} P_i = \sum_{j(j \neq i)} (\Gamma_{i,j} P_j - \Gamma_{j,i} P_i) \quad \sum_i P_i = 1 \quad (13)$$

where $\Gamma_{i,j}$ are the elements of the matrix of the transition rates, $\Gamma = \Gamma^{l,+} + \Gamma^{r,+} + \Gamma^{l,-} + \Gamma^{r,-}$. From the stationary solution $\{\bar{P}_i\}$, which is obtained by putting $dP_i/dt = 0$ in (13) the DC current, the number of electrons that pass the left/right barrier per unit of time, is determined

$$I \equiv I^{l/r} = \mp e \sum_{i,j(j \neq i)} \bar{P}_j (\Gamma_{i,j}^{l/r,-} - \Gamma_{i,j}^{l/r,+}).$$

† It is well known that strong dissipation can suppress tunnelling [17]. This choice for the phase-breaking rate τ_Φ^{-1} guarantees that the renormalization of the tunnelling rates through the barriers is negligible.

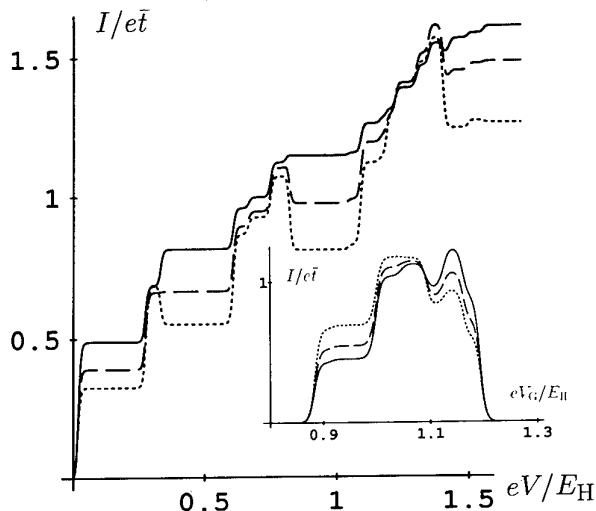


Figure 6. Current–voltage characteristics ($\mu_r = 0$) and the splitting of the fourth conductance peak at $\mu_1 = 0.3E_H$ and $\mu_r = 0$ (inset) of a dot of length $L = 15a_B$ described by the correlated electron model for $\beta = 200/E_H$. Tunnelling integrals are $t_2 = 0.03E_H$, $t_3 = 0.07E_H$ and $t_4 = 0.09E_H$, numerically determined ground state energies $E_0(1) = 0.023E_H$, $E_0(2) = 0.30E_H$, $E_0(3) = 0.97E_H$, $E_0(4) = 2.15E_H$. Broken, dotted and full curves correspond to $t'/t^r = 1, 2$ and 0.5 respectively. The current is plotted in units of the total transmission rate $\bar{t} = t't^r/(t^l + t^r)$.

Current–voltage characteristics and conductivity peaks calculated by using the excitation energies given in table 1 are shown in figure 6 for temperatures lower than the excitation energies. We observe fine structure in the Coulomb staircase consistent with recent experiments [8, 9], and earlier theoretical predictions using a different approach [18, 19]. Within our model, the Coulomb steps are not of equal length as in the phenomenological charging model used previously [18]. This deviation from the classical behaviour is related to the inhomogeneity of the quantum mechanical charge density of the ground state [4, 15]. The heights of the fine structure steps are more random due to the non-regular sequence of total spins (cf table 1) and the spin selection rules. In certain cases, fine structure steps in the I – V characteristic may even be completely suppressed.

Strikingly, regions of negative differential conductance occur (figures 6 and 7). They are related to the reduced probability for the states with maximal spin, $S = N/2$, to decay into states with lower electron number. In contrast to transitions that involve $S < N/2$ they are only possible if S is reduced. The corresponding Clebsch–Gordan coefficients are smaller than those for transitions with increasing S (cf equation (12)) which leads to an additional reduction of the current as compared with the situation where $S < N/2$. When the voltage is raised to such a value that an $S = N/2$ state becomes involved in the transport, this state attracts considerable stationary population at the expense of the better conducting $S < N/2$ states, as can be seen in figure 8.[†] Both effects

[†]The populations shown here do not sum up to unity because of the occupation of other states.

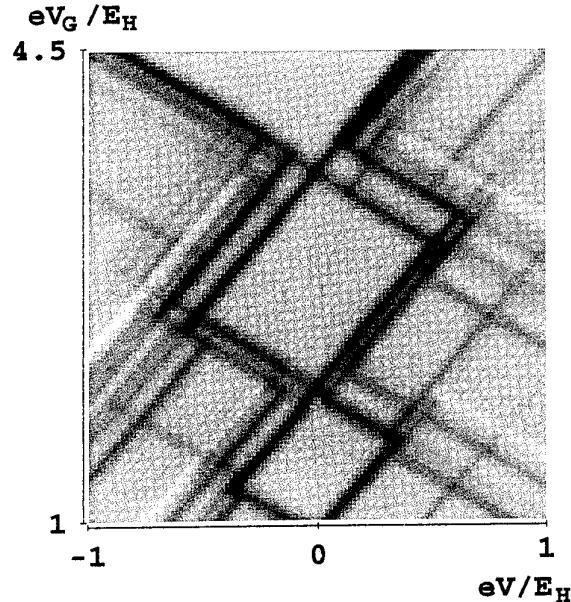


Figure 7. Grey scale plot of the differential conductance as a function of the gate voltage and the transport voltage. Negative differential conductances are indicated by light regions.

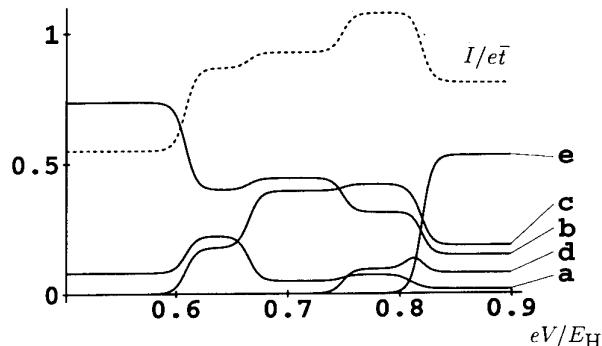


Figure 8. The most prominent feature in figure 6 for $t^l > t^r$ is magnified (dotted curve, in units of $e\bar{t}$), and the corresponding populations of the most relevant dot states **a**: $N = 2, S = 0$, **b**: $N = 2, S = 1$, **c**: $N = 3, S = 1/2$ (ground state), **d**: $N = 3, S = 1/2$ (first excited state), **e**: $N = 3, S = 3/2$ versus bias voltage V . Decreasing current is accompanied by an increase of the population of the spin polarized $N = 3, S = 3/2$ state at the expense of the other populations.

together can then add up to a decreasing current. The decrease in the I – V curve becomes less pronounced if $t^l < t^r$, because then the dot is mostly empty and the $(N - 1) \rightarrow N$ transitions determine the current. On the other hand if $t^l > t^r$ the *spin blockade* becomes more pronounced, because the $N \rightarrow (N - 1)$ transitions limit the current in this case. Both features can be observed in figure 6 and also the experimental data of [8, 9] are clearly consistent with an interpretation that the potential barriers are slightly different. Negative differential conductances can in principle be used to construct a mesoscopic oscillator.

For low V , the conductance shows peaks when V_G is varied, which can be described within thermal equilibrium in the limit of linear response [21]. Only the (correlated N -electron) ground states are involved at zero temperature. For finite bias voltages, $eV = \mu_l - \mu_r$, larger than the level spacing, a varying number of levels contribute to the current when V_G is changed. The conductance peaks split and show structure as is observed experimentally and explained qualitatively in [8, 22] within the charging model. Asymmetric coupling to the leads changes the shape of the peaks considerably, as can be seen in figure 6. We propose to explain the ‘inclination’ of the conductance peaks observed in the experiment [8, 9] by asymmetric barriers and predict that this inclination will be reversed if the sign of the bias voltage is changed. Such asymmetric conductance properties can be used to construct mesoscopic rectifiers. Similar effects were inferred earlier from the high-frequency properties of mesoscopic systems containing asymmetric disorder [23].

5. Conclusions

In summary, we predict a reduction of the conductance plateaus for quantum wires due to an onset of acoustical phonon scattering at low temperatures for $v_F > c_s$ well below the onset of a new conductance plateau. The underlying physics can be understood using energy and momentum conservation. It should be observable in a characteristic decrease of the conductance at Fermi energy and/or magnetic fields that correspond to $v_F > c_s$ when the temperature is increased. This is in contrast to the temperature-dependent smearing of disorder-induced anti-resonances at the onset of a new plateau, which would lead to an increase of the conductance with increasing temperature.

Furthermore, it is shown that correlations in semiconducting quantum dots qualitatively influence the spectrum and its variation with dot diameter. The lowest excitation energies involve spin and can be understood through the inhomogeneous charge density distribution. Their ratios are not sensitive to the form of the $e^- - e^-$ repulsion.

The Coulomb and spin effects lead to conventional Coulomb blockade and a novel spin blockade effect in nonlinear transport through a double barrier. Most strikingly, regions of negative differential conductance occur because for each electron number the state of maximum spin can only contribute to transport by reducing the total spin. As a consequence, the transition probability into states with lower electron number is reduced. Spin blockade is not restricted to the quasi-1D model considered here but should also occur in 2D systems used in the experiments as long as the density of the electrons is sufficiently small.

All of the theoretically predicted features described above are qualitatively consistent with experiment [8, 9]. Further experiments, in particular using ‘slim quantum

dots’, are, however, necessary to clarify the quantitative aspects.

‘Preliminary’ results taking into account a magnetic field in the direction of the transport show that the negative differential conductance is influenced and suppressed at high fields. To clarify these questions and to be able to make quantitative comparison with existing experiments data, generalization of the above correlated electron model to 2D is necessary.

Acknowledgments

We thank W Apel, R Haug, J Weis and U Weiss for valuable discussions. We are also very grateful to J Wróbel and F Kuchar for discussions about their experiments. This work was supported in part by the Deutsche Forschungsgemeinschaft via grants We 1124/2-2, We 1124/4-1 and AP 47/1-1 and by the European Community within the SCIENCE program, grant SCC*-CT90-0020.

References

- [1] For a general discussion of dephasing and inelastic scattering, see Stern A, Aharonov Y and Imry Y 1990 *Phys. Rev. A* **41** 3436
Imry Y and Stern A 1994 *Semicond. Sci. Technol.* **9** 1879
- [2] Brandes T and Kramer B 1993 *Solid State Commun.* **88** 773
- [3] Meirav U, Kastner M A and Wind S J 1990 *Phys. Rev. Lett.* **65** 771
Kastner M A 1992 *Rev. Mod. Phys.* **64** 849
- [4] Jauregui K, Häusler W and Kramer B 1993 *Europhys. Lett.* **24** 581
Brandes T, Jauregui K, Häusler W, Kramer B and Weinmann D 1993 *Physica B* **189** 16
- [5] Häusler W and Kramer B 1993 *Phys. Rev. B* **47** 16353
- [6] Häusler W, Jauregui K, Weinmann D, Brandes T and Kramer B 1994 *Physica B* **194–196** 1325
- [7] Weinmann D, Häusler W, Pfaff W, Kramer B and Weiss U 1994 *Europhys. Lett.* **26** 467
- [8] Johnson A T, Kouwenhoven L P, de Jong W, van der Vaart N C, Harmans C J P M and Foxon C T 1992 *Phys. Rev. Lett.* **69** 1592
- [9] Weis J, Haug R J, von Klitzing K and Ploog K 1993 *Phys. Rev. Lett.* **71** 4019; 1992 *Phys. Rev. B* **46** 12837
- [10] Heinonen O, Taylor P L and Girvin S M 1984 *Phys. Rev. B* **30** 3016
- [11] Středa P and von Klitzing K 1984 *J. Phys. C: Solid State Phys.* **17** L483
- [12] Landauer R 1970 *Phil. Mag.* **21** 863
Büttiker M 1986 *Phys. Rev. Lett.* **57** 1761
- [13] Wróbel J, Kuchar F, Ismail K, Lee K Y, Nickel H, Schlapp W, Grabecki G and Dietl T 1993 *10th Int. Conf. on Electronic Properties of Two-Dimensional Systems* (Newport, 1993)
- [14] Kinaret J M and Lee P A 1990 *Phys. Rev. B* **42** 11768
- [15] Häusler W, Kramer B and Mašek J 1991 *Z. Phys. B* **85** 435
- [16] Häusler W 1994 *Preprint*
- [17] Weiss U 1993 *Quantum Dissipative Systems (Series in Modern Condensed Matter Physics, vol 2)* (Singapore: World Scientific)

- [18] Averin D V and Korotkov A N 1990 *J. Low Temp. Phys.* **80** 173
- [19] Averin D V, Korotkov A N and Likharev K K 1991 *Phys. Rev. B* **44** 6199
- [20] Beenakker C W J 1991 *Phys. Rev. B* **44** 1646
- [21] Meir Y, Wingreen N S and Lee P A 1991 *Phys. Rev. Lett.* **66** 3048
- [22] Foxman E B, McEuen P L, Meirav U, Wingreen N S, Meir Y, Belk P A, Belk N R, Kastner M A and Wind S J 1993 *Phys. Rev. B* **47** 10020
- [23] Fal'ko V I 1989 *Europhys. Lett.* **8** 785

Dephasing by coupling with the environment, application to Coulomb electron-electron interactions in metals

Yoseph Imry[†] and Ady Stern[‡]

[†] Department of Condensed-Matter Physics, Weizmann Institute, 76100 Rehovot, Israel

[‡] Physics Department, Harvard University, Cambridge, MA 02138, USA

Abstract. A general formulation will be given of the loss of phase coherence between two partial waves, leading to the *dephasing* of their interference. This is due to inelastic scattering from the ‘environment’ (which is a different set of degrees of freedom than the waves are coupled with). For a conduction electron, the other electrons (‘Fermi sea’) are often the dominant environment of this type. Coulomb interactions with the latter are, especially at lower dimensions, the most important dephasing mechanism. It will be shown how this picture yields rather straightforwardly the very non-trivial results of Altshuler, Aronov and Khmelnitskii in one and two dimensions, in the diffusive case. Subtleties associated with divergences that have to be subtracted will be discussed. These results are known to agree well with experiments.

As a new application of the above ideas, the dephasing in a zero-dimensional quantum dot will be briefly considered. This will lead to stringent conditions for observing the discrete spectrum of such a dot, in agreement with recent experiments. The crossover at low temperatures in small wires from one- to zero-dimensional behaviour will be shown to ‘rescue’ the Landau Fermi-liquid theory from being violated because of the $T^{2/3}$ behaviour of the 1D dephasing rate.

After clarifying the relationship between the e-e scattering rate and the dephasing rate, the connection with the former will be made, including the ballistic regime.

1. Introduction and review of the principles of dephasing

Many of the interesting effects in mesoscopic systems are due to quantum interference. Among these are, for example, the weak localization corrections to the conductivity, the universal conductance fluctuations, persistent currents and many others. These effects are known to be affected by the coupling of the interfering particle to its environment, for example, to a heat bath. The way in which such a coupling modifies quantum phenomena has been studied for a long time, both theoretically (Feynman and Vernon 1963, Caldeira and Leggett 1983) and experimentally. The effect of the coupling to the environment may be characterized by the ‘phase-breaking’ time, τ_ϕ , which is the characteristic time for the interfering particle to stay phase coherent, as explained below.

Stern *et al* (1990a,b) studied the way in which coupling of an interfering particle affects a two-wave interference experiment. Our discussion in the first three sections will be mostly based on that work. Two descriptions have been used for the way in which the interaction of a quantum system with its environment might suppress quantum interference. The first regards

the environment as measuring the path of the interfering particle. When the environment has the information on that path, no interference is seen. The second description answers the question naturally raised by the first: how does the interfering particle ‘know’, when the interference is examined, that the environment has identified its path? This question is answered by the observation that the interaction of a partial wave with its environment can induce an uncertainty in this wave’s phase (what counts physically is the uncertainty of the *relative* phases of the paths). This can be described as turning the interference pattern into a sum of many patterns, shifted relative to one another. The two descriptions were proved to be equivalent, and this has been applied to the dephasing by electromagnetic fluctuations in metals, and by photon modes in thermal and coherent states. Here we will review the two descriptions, and examine in sections 2–4 the dephasing by the electron–electron interaction in metals. We shall find it convenient to consider that problem from the first point of view mentioned above, rather than the second; that is, to find out where the information on the interfering electron path is hidden in the bath of electrons with which it interacts.

As a guiding example, we consider an Aharonov–Bohm (A–B) interference experiment on a ring. The A–B

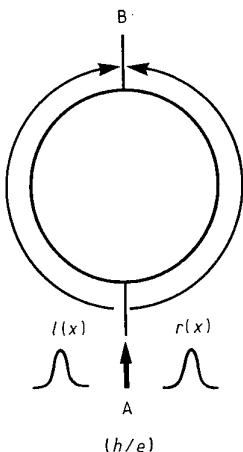


Figure 1. Schematics of interference experiments in A-B rings. Each partial wave traverses half the ring, and the interference is examined at the point B. This kind of interference gives rise to h/e oscillations of the conductance.

effect has been proved to be a convenient way to observe interference patterns in mesoscopic samples, because it provides an experimentally straightforward way of shifting the interference pattern. This experiment starts by a construction of two electron wave packets, $l(x)$ and $r(x)$ (l, r stand for left, right), crossing the ring (see figure 1) along its two opposite sides. We assume that the two wave packets follow well defined classical paths, $x_l(t)$, $x_r(t)$ along the arms of the ring. The interference is examined after each of the two wave packets have traversed half of the ring's circumference. Therefore, the initial wavefunction of the electron (whose coordinate is x) and the environment (whose wavefunction and set of coordinates are respectively denoted by χ and η) is

$$\psi(t=0) = [l(x) + r(x)] \otimes \chi_0(\eta). \quad (1)$$

At time τ_0 , when the interference is examined, the wavefunction is, in general

$$\psi(\tau_0) = l(x, \tau_0) \otimes \chi_l(\eta, \tau_0) + r(x, \tau_0) \otimes \chi_r(\eta, \tau_0) \quad (2)$$

and the interference term is (see below)

$$2 \operatorname{Re} \left(l^*(x, \tau_0) r(x, \tau_0) \int d\eta \chi_l^*(\eta, \tau_0) \chi_r(\eta, \tau_0) \right). \quad (3)$$

Had there been no environment present in the experiment, the interference term would have been just $2 \operatorname{Re}(l^*(x, \tau_0) r(x, \tau_0))$. So, the effect of the interaction is to multiply the interference term $\int d\eta \chi_l^*(\eta, \tau_0) \chi_r(\eta, \tau_0)$ at τ_0 . This is so since the environment is not observed in the interference experiment; its coordinate is therefore integrated upon, that is, the scalar product of the two environmental states at τ_0 is taken. The first way to understand the dephasing is seen directly from this expression, which is the scalar product of the two environment states at τ_0 , coupled to the two partial waves. At $t = 0$ these two states are identical. During the time of the experiment, each partial wave has its own interaction with the environment, and therefore the two states evolving in time become different. When the two

states of the environment become orthogonal, the final state of the environment identifies the path the electron took. Quantum interference, which is the result of an uncertainty in this path, is then lost. Thus, the phase-breaking time, τ_ϕ , is the time in which the two interfering partial waves shift the environment into states orthogonal to each other, that is when the environment has the information on the path that the electron takes.

The second explanation for the loss of quantum interference regards it from the point of view of how the environment affects the partial waves, rather than how the waves affect the environment. It is well known that when a static potential $V(x)$ is exerted on one of the partial waves, this wave accumulates a phase

$$\phi = - \int V(x(t)) dt / \hbar \quad (4)$$

and the interference term is multiplied by $e^{i\phi}$. ‘A static potential’ here is a potential which is a function of the particle’s coordinate and momentum only, and does not involve any other degrees of freedom. For a given particle’s path, the value of a static potential is well defined. When V is not static, but created by the degree(s) of freedom of the environment, V becomes an operator. Thus its value is no longer well defined. The uncertainty in this value results from the quantum uncertainty in the state of the environment. Therefore, ϕ is not definite either. In fact, ϕ becomes a statistical variable, described by a distribution function $P(\phi)$. (For the details of this description see Stern *et al* (1990a,b).) The effect of the environment on the interference is then to multiply the interference term by the average value of $e^{i\phi}$, i.e.

$$\langle e^{i\phi} \rangle = \int P(\phi) e^{i\phi} d\phi. \quad (5)$$

The averaging is done on the interference ‘screen’. Since $e^{i\phi}$ is periodic in ϕ , $\langle e^{i\phi} \rangle$ tends to zero when $P(\phi)$ is slowly varying over a region much larger than one period, of 2π . When this happens, one may say that the interference screen shows a superposition of many interference patterns, mutually cancelling each other. Hence, the phase-breaking time is also the time in which the uncertainty in the phase becomes of the order of the interference periodicity. In the Feynman–Vernon terminology, $\langle e^{i\phi} \rangle$ is the influence functional of the two paths taken by the two partial waves. This is, then, the second explanation for the loss of quantum interference.

The statement of equivalence between the two explanations is given by the equation

$$\langle e^{i\phi} \rangle = \int d\eta \chi_l^*(\eta) \chi_r(\eta). \quad (6)$$

When the environment measures the path taken by the particle (by χ_l becoming orthogonal to χ_r), it induces a phaseshift whose uncertainty is of the order of 2π . The equivalence embodied in equation (6) is proved as follows.

We start by considering dephasing of the right-hand path χ_r only. The generalization to two paths will be seen

later. The Hamiltonian of the environment will be denoted by $H_{\text{env}}(\eta, p_\eta)$, while the interaction term is $V(\chi_r(t), \eta)$ (the left partial wave does not interact with the environment). Starting with the initial wavefunction (1) the wavefunction at time τ_0 is

$$\begin{aligned}\psi(\tau_0) &= l(\tau_0) \exp(-iH_{\text{env}}\tau_0/\hbar)\chi_0(\eta) \\ &+ r(\tau_0)\hat{T} \exp\left(-\frac{i}{\hbar} \int_0^{\tau_0} dt (H_{\text{env}} + V)\right)\chi_0(\eta)\end{aligned}\quad (7)$$

where \hat{T} is the time-ordering operator. It is useful at this point to write $\psi(\tau_0)$ in terms of

$$V_l(t) \equiv \exp(iH_{\text{env}}t)V(\chi_r(t), \eta)\exp(-iH_{\text{env}}t),$$

i.e. the potential V in the interaction picture. Using V_l , $\psi(\tau_0)$ can be written as

$$\begin{aligned}\psi(\tau_0) &\equiv l(\tau_0) \otimes \exp(-iH_{\text{env}}\tau_0/\hbar)\chi_0(\eta) \\ &+ r(\tau_0) \otimes \exp(-iH_{\text{env}}\tau_0/\hbar)\hat{T} \\ &\times \exp\left(-i \int_0^{\tau_0} \frac{dt}{\hbar} V_l(x_r(t), t)\right)\chi_0(\eta).\end{aligned}\quad (8)$$

Hence the interference term is multiplied by

$$\begin{aligned}\langle \chi_0 | \exp(iH_{\text{env}}\tau_0)\hat{T} \exp\left(-i \int_0^{\tau_0} \frac{dt}{\hbar} (H_{\text{env}} + V)\right) | \chi_0 \rangle \\ = \langle \chi_0 | \hat{T} \exp\left(-\frac{i}{\hbar} \int_0^{\tau_0} dt V_l(x_r(t), t)\right) | \chi_0 \rangle.\end{aligned}\quad (9)$$

The interpretation of this expression in terms of a scalar product of two environment states at time τ_0 is obvious. The interpretation in terms of phase uncertainty emerges from the observation that equation (9) is the expectation value of a unitary operator. As all unitary operators, this operator can be expressed as the exponent of a Hermitian operator ϕ , i.e.

$$\langle \chi_0 | \hat{T} \exp\left(-\frac{i}{\hbar} \int_0^{\tau_0} dt V_l(x_r(t), t)\right) | \chi_0 \rangle = \langle \chi_0 | e^{i\phi} | \chi_0 \rangle. \quad (10)$$

Hence the effect of the interaction with the environment is to multiply the interference term by $\langle e^{i\phi} \rangle$, where the averaging is done with respect to the phase probability distribution, as determined by the environmental state χ_0 .

The phase operator ϕ was introduced here by means of the mathematical properties of unitary transformations, so that it still deserves a physical interpretation. To obtain such an interpretation, we first discuss the case where the potentials exerted by the environment at different points along the particle's path commute, i.e.

$$[V_l(x_r(t), t), V_l(x_r(t'), t')] = 0 \quad (11)$$

Then

$$\begin{aligned}\langle \chi_0 | \hat{T} \exp\left(-\frac{i}{\hbar} \int_0^{\tau_0} dt V_l(x_r(t), t)\right) | \chi_0 \rangle \\ = \langle \chi_0 | \exp\left(-\frac{i}{\hbar} \int_0^{\tau_0} dt V_l(x_r(t), t)\right) | \chi_0 \rangle\end{aligned}\quad (12)$$

and

$$\phi = -\frac{1}{\hbar} \int_0^{\tau_0} dt V_l(x_r(t), t).$$

In this case $\dot{\phi}$, the rate of accumulation of the phase, is just the local potential acting on the interfering particle, independent of earlier interactions of the particle with the environment. One should distinguish here between two limits: for $\langle \delta\phi^2 \rangle \ll 1$, equation (10) yields

$$\langle e^{i\phi} \rangle \approx e^{i\langle \phi \rangle}(1 - \frac{1}{2}\langle \delta\phi^2 \rangle) \quad (13)$$

and the environment's potential can be approximated by a single-particle (possibly time-dependent) potential

$$\langle V_l(x_r(t), t) \rangle = \langle \chi_0 | V_l(x_r(t), t) | \chi_0 \rangle. \quad (14)$$

For $\langle \delta\phi^2 \rangle \gg 1$, on the other hand, the interference term tends to zero. The crossover between the two regimes is then at

$$\begin{aligned}\langle \delta\phi^2 \rangle &= \int_0^{\tau_0} \frac{dt}{\hbar} \int_0^{\tau_0} \frac{dt'}{\hbar} [\langle V_l(x_r(t), t) V_l(x_r(t'), t') \rangle \\ &- \langle V_l(x_r(t), t) \rangle \langle V_l(x_r(t'), t') \rangle] \\ &= \int_0^{\tau_0} \frac{dt}{\hbar} \int_0^{\tau_0} \frac{dt'}{\hbar} \left(\int d\eta \int d\eta' \chi^*(\eta', t') V(x_r(t'), \eta') \right. \\ &\times \langle \eta' | \exp[-iH_{\text{env}}(t' - t)] | \eta \rangle \\ &\times V(x_r(t), \eta) \chi(\eta, t) \\ &- \int d\eta \chi^*(\eta, t) V(x_r(t), \eta) \chi(\eta, t) \\ &\left. \times \int d\eta' \chi^*(\eta', t') V(x_r(t'), \eta') \chi(\eta', t') \right) \sim 1 \quad (15)\end{aligned}$$

where $\chi_0(\eta, t) \equiv \exp(-iH_{\text{env}}t)\chi_0(\eta)$ is the environmental state as it evolves in time under H_{env} .

When is the condition in equation (11) valid, and what happens when it is not? A typical case where the potentials at different points along the path are commutative is the case of an interfering electron interacting with a free electromagnetic field. In that case the interaction is

$$V_l(x_r(t), t) = -\frac{e}{c} \dot{x}_r(t) \cdot A(x_r(t), t) \quad (16)$$

where $A(x, t)$, the electromagnetic free field, is in obvious notation

$$A(x, t) = \sum_{k, \lambda} \varepsilon_{k, \lambda} \left(\frac{2\pi c^2}{\omega_k} \right)^{1/2} (a_k e^{ik \cdot x - i\omega t} + a_k^\dagger e^{-ik \cdot x + i\omega t}) \quad (17)$$

and $[V_l(x, t), V_l(x', t')] = 0$ unless $|x - x'| = c|t - t'|$. Since $\dot{x}_r(t) < c$, condition (11) is valid. Generally, this condition is valid when there is no amplitude for an environmental excitation created at $(x_r(t), t)$ to be annihilated at $(x_r(t'), t')$, i.e. when a change induced in the environment's state at $(x_r(t), t)$ does not influence the potential the environment exerts on the interfering

particle at $(x_r(t'), t')$. In the above example, a photon emitted by the electron at $(x_r(t), t)$ will not be at $(x_r(t'), t')$ when the electron gets there.

If, instead of discussing the interaction with a photon field, we turn our attention to the interaction with phonons, the speed of light in equations (6) and (17) is replaced by the sound velocity in the analogous theory. Then, a phonon emitted by the electron at $(x_r(t), t)$ might be encountered again by the electron at $(x_r(t'), t')$. Hence lattice excitations excited by the electron along its path may affect the potential it feels at a later stage of the path. The potential the electron feels at a given point of its path is now not a local function of that point, but depends on the path since it includes a ‘back reaction’ of the environment to the potential exerted by the electron. Therefore this potential will be different from $V_l(x_r(t), t)$, and consequently, the rate of phase accumulation will also differ from $V_l(x_r(t), t)$. However, in large many-body environments the potential exerted by the environment on the interfering particle is usually practically independent of the particle’s history since the environment’s memory time is very short. Therefore equation (11) can be assumed to hold.

We thus see that the loss of interference due to an interaction with a dynamical environment can be understood in the two ways discussed. The interference is destroyed either when the state of the environment coupled to the right wave is orthogonal to that coupled to the left wave, or, alternatively, when the width of the phase distribution function exceeds a magnitude of order unity. The interaction with the dynamical environment turns the phase into a statistical variable, and this, together with the phase being defined only over a range of 2π , determines the conditions for the phase to become completely uncertain. If the potential exerted by the environment on the interfering particle at a given point along its path is assumed to be independent of the path, the phase uncertainty is given by,

$$\langle \delta\phi^2 \rangle = \int_0^{t_0} \frac{dt}{\hbar} \int_0^{t_0} \frac{dt'}{\hbar} [\langle V_l(x_r(t), t) V_l(x_r(t'), t') \rangle - \langle V_l(x_r(t), t) \rangle \times \langle V_l(x_r(t'), t') \rangle]. \quad (18)$$

The exact behaviour of the interference term for $\langle \delta\phi^2 \rangle \gg 1$, i.e. the value of $\langle e^{i\phi} \rangle$ for broad distribution functions, depends on the phase distribution, $P(\phi)$. However, the description of the phase as a statistical variable enables us, under appropriate conditions, to apply the central limit theorem, and conclude that $P(\phi)$ is a normal distribution. The central limit theorem is applicable, for example, when the phase is accumulated in a series of uncorrelated events (for example by a series of scattering events off different, non-interacting scatterers), or, more generally, whenever the potential–potential correlation function decays to zero with a characteristic decay time much shorter than the duration of the experiment. In particular, the central limit theorem is usually applicable for coupling to a heat bath. For a normal distribution

$$\langle e^{i\phi} \rangle = \exp[i\langle \phi \rangle - \frac{1}{2}\langle \delta\phi^2 \rangle]. \quad (19)$$

This expression is exact for the model of an environment composed of harmonic oscillators with a linear coupling to the interfering waves. The evaluation of $\langle e^{i\phi} \rangle$ by equation (19) reproduces the result obtained by Feynman and Vernon for a rather similar model. Feynman and Vernon’s result was obtained by integration of the environment’s paths. This model has proved in recent years to be very useful in the investigation of the effect of the environment on quantum phenomena (for example, Caldeira and Leggett 1983). Equation (19) is therefore a convenient way to calculate the influence functional for many-body environments, where the central limit theorem is usually applicable.

As seen from equation (15), the phase uncertainty remains constant when the interfering wave does not interact with the environment. Thus, if a trace is left by a partial wave on its environment, this trace cannot be wiped out after the interaction is over. Neither internal interactions of the environment nor a deliberate application of a classical force on it can reduce the phase uncertainty after the interaction with the environment is over. This statement can also be proved from the point of view of the change that the interfering wave induces in its environment. This proof follows simply from unitarity. The scalar product of two states that evolve in time under the same Hamiltonian does not change in time. Therefore, if the state of the system (electron plus environment) after the electron–environment interaction took place is

$$|r(t)\rangle \otimes |\chi_{\text{env}}^{(1)}\rangle + |l(t)\rangle \otimes |\chi_{\text{env}}^{(2)}\rangle \quad (20)$$

then the scalar product $\langle \chi_{\text{env}}^{(1)}(t) | \chi_{\text{env}}^{(2)}(t) \rangle$ does not change in time. The only way to change it is by another interaction of the electron with the same environment. Such an interaction keeps the product $\langle \chi_{\text{env}}^{(1)}(t) | \chi_{\text{env}}^{(2)}(t) \rangle \otimes \langle r(t) | l(t) \rangle$ constant, but changes $\langle \chi_{\text{env}}^{(1)}(t) | \chi_{\text{env}}^{(2)}(t) \rangle$. The interference will be retrieved only if the orthogonality is transferred from the environment wavefunction to the electronic wavefunctions which are not traced on in the experiment.

The above discussion was concerned with the phase $\phi = \phi_r$, accumulated by the right-hand path only. The left-hand path accumulates similarly a phase ϕ_l from the interaction with the environment. The interference pattern is governed by the relative phase $\phi_r - \phi_l$, and it is the uncertainty in that phase which determines the loss of quantum interference. This uncertainty is always smaller than, or equal to, the sum of uncertainties in the two partial waves’ phases. The case of non-commuting phases will not be discussed here.

Often the same environment interacts with the two interfering waves. A typical example is the interaction of an interfering electron with the electromagnetic fluctuations in a vacuum. In this case, if the two waves follow parallel paths with equal velocities, their dipole radiation, despite the energy it transfers to the field, does not dephase the interference. This radiation makes each of the partial waves’ phases uncertain, but does not alter the relative phase. We shall encounter more examples later demonstrating that the environmental excitation

created must be able to distinguish l from r in order to dephase their interference.

The last example demonstrates that an exchange of energy is not a sufficient condition for dephasing. It is also not a necessary condition for dephasing. What is important is that the two partial waves flip the environment to *orthogonal* states. It does not matter in principle that these states are degenerate. Simple examples were given by Stern *et al* (1990a). Thus, it must be emphasized that, for example, long-wave excitations (phonons, photons) may not dephase the interference. But that is *not* because of their low energy but rather because they may not influence the *relative* phase of the paths.

We emphasize that dephasing may occur by coupling to a discrete or a continuous environment. In the former case the interfering particle is more likely to ‘reabsorb’ the excitation and ‘reset’ the phase. In the latter case, practically speaking, the excitation may move away to infinity and the loss of phase can usually be regarded as irreversible. The latter case is that of an effective ‘bath’ and there are no subtleties with the definition of ϕ since equation (11) may be assumed. We point out that in special cases it is possible, even in the continuum case, to have a finite probability to reabsorb the created excitation and thus retain coherence. This happens, for example, in a quantum interference model due to Holstein (1961) for the Hall effect in insulators.

2. Dephasing by the electron-electron interaction

An interesting application of the above general principle is the dephasing the mesoscopic interference effects by electron-electron interaction in conducting samples. Stern *et al* (1990a,b) have applied the phase uncertainty approach to dephasing by electron-electron interaction in metals in the diffusive regime. They have shown that this approach reproduces the results obtained in the pioneering work of Alt'shuler *et al* (1982). Following Stern *et al* (1990b) we now consider the dephasing due to electron-electron interactions from the point of view of the changes induced in the state of the environment, using the response functions of the latter. In the original work of Altshuler *et al* the phase uncertainty induced on the particle by the electromagnetic fluctuations of the environment was considered. The fluctuation-dissipation theorem guarantees the equivalence of these two pictures.

The general picture is that of a test particle interacting with an environment. For definiteness, we consider an interfering ‘electron’, whose paths are denoted by $x_{r,l}(t)$, interacting with a bath of environment electrons, whose coordinates are y_i . The identity of the interfering electron with those of the bath will be approximately handled later. The Coulomb interaction of the interfering electron with the rest of the electrons is, in the interaction picture

$$\hat{V}_l(\mathbf{x}, t) = \int \frac{\hat{\rho}_l(\mathbf{r}', t) d^3 r'}{|\mathbf{x} - \mathbf{r}'|} \quad (21)$$

where $\hat{\rho}_l(r, t) = e \sum_i \delta(r - \hat{y}_i^l(t))$. For the brevity of the

following expressions, we consider only the interaction of the electron bath with the right partial wave of the interfering electron, we omit the corresponding subscript, and we assume first (this will be relaxed later) that the electron bath is initially in its ground state, $|0\rangle$. Assuming that the left partial wave does not interact with the electron bath, the intensity of the interference pattern is reduced by the probability that the bath's state coupled to the right wave becomes different from $|0\rangle$. Up to second order in the interaction, this probability is

$$P = \frac{1}{\hbar^2} \sum_{|n\rangle \neq |0\rangle} \int_0^{t_0} dt \times \int_0^{t_0} dt' \langle 0 | V_l(\mathbf{x}(t), t) | n \rangle \langle n | V_l(\mathbf{x}(t'), t') | 0 \rangle. \quad (22)$$

For the ground state $\langle 0 | \hat{\rho} | 0 \rangle = 0$, so that the summation in equation (22) can be extended to include all states. We neglect the changes in the paths $x_{r,l}(t)$ due to the interaction; only the phase due to the latter is taken into account. The interpretation of equation (22) as the variance of the phase given to the particle by the interaction with the environment is clear. We now express P in terms of the response of the environment. Using the convolution theorem,

$$\int d^3 r' f(\mathbf{r} - \mathbf{r}') g(\mathbf{r}') = \frac{1}{(2\pi)^3} \int d^3 q f_q g_q e^{-i\mathbf{q} \cdot \mathbf{r}}$$

where f_q and g_q are the Fourier transforms of f and g , we write

$$P = \frac{1}{\hbar^2 (2\pi)^6} \int_0^{t_0} dt \int_0^{t_0} dt' \int d^3 q \times \int d^3 q' \frac{4\pi e}{q^2} \frac{4\pi e}{q'^2} \langle \rho_q(t) \rho_{q'}(t') \rangle \exp(i\mathbf{q} \cdot \mathbf{x}(t) - i\mathbf{q}' \cdot \mathbf{x}(t')). \quad (23)$$

We assume translational invariance

$$\langle \rho_q \rho_{q'} \rangle = \frac{(2\pi)^3}{Vol} \delta(\mathbf{q} + \mathbf{q}') \langle \rho_q \rho_{-\mathbf{q}} \rangle. \quad (24)$$

(For a finite system the q 's are discrete and one just has $\delta_{qq'}$. Going to the continuum the Kronecker delta is replaced by $(2\pi)^3/Vol$ times the Dirac delta.) Performing one q integration and inserting a complete set of intermediate states, we obtain

$$P = \frac{1}{Vol(2\pi)^3 \hbar^2} \sum_{|n\rangle} \int_0^{t_0} dt \int_0^{t_0} dt' \int d^3 q \frac{(4\pi e)^2}{q^4} \times \langle 0 | \rho_l^q(t) | n \rangle \langle n | \rho_l^q(t') | 0 \rangle \exp[i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(t'))]. \quad (25)$$

By transforming into Schrödinger picture operators and inserting a dummy integration variable ω , P can be

rewritten in the following form

$$\begin{aligned} P = & \frac{1}{\text{Vol}(2\pi)^3 \hbar^2} \sum_{|n\rangle} \int_0^{\tau_0} dt \int_0^{\tau_0} dt' \int d^3 q \\ & \times \int d\omega \frac{(4\pi e)^2}{q^4} |\langle 0 | \rho_S^q | n \rangle|^2 \delta(\omega - \omega_{n0}) \\ & \times \exp[i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(t')) - i\omega(t - t')]. \end{aligned} \quad (26)$$

At a first glance, equation (26) looks useless, due to the practical impossibility of calculating the bath's eigenstates $|n\rangle$. However, the usefulness of equation (26) stems from its relation to the linear response expression for dynamic structure factor and the imaginary part of the complex dielectric function where both are related by the fluctuation-dissipation theorem (F-D) (see for example Nozières 1964)

$$\begin{aligned} \text{Im}\left(\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right) &= \frac{4\pi^2 e^2}{\text{Vol} q^2 \hbar} \sum_{|n\rangle} |\langle 0 | \rho_S^q | n \rangle|^2 \delta(\omega - \omega_{n0}) \\ &= \frac{4\pi^2 e^2}{q^2} S(\mathbf{q}, \omega). \end{aligned} \quad (27)$$

Thus, equation (26) becomes

$$\begin{aligned} P = & \frac{1}{\hbar(2\pi)^3} \int_0^{\tau_0} dt \int_0^{\tau_0} dt' \int d^3 q \int d\omega \frac{4e^2}{q^2} \text{Im}\left(\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right) \\ & \times \exp[i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(t')) - i\omega(t - t')]. \end{aligned} \quad (28)$$

Equation (28) is the powerful central result of this section. Before proceeding to a discussion of this result, we comment that the calculation can be generalized to treat an electron bath initially in a thermal state. The integrand in equation (28) is then multiplied by $\coth(\omega/2k_B T)$. The probability that the state of the environment will change during the time τ_0 is expressed through an integral over the *dissipative* part of the response, i.e. the excitability of the system. The equivalent expression via the dynamic structure factor through the F-D theorem simply expresses this by integrals over the inelastic scattering probability. It is well known (see, for example, Nozières 1964) that the energy loss to the system by inelastic scattering is given in a similar fashion by an integral over S or $\text{Im}(1/\varepsilon)$. The new feature of our result is the appearance of the classical path $\mathbf{x}(t)$ along which the excitation of the environment occurs. The phase in the exponentials in equation (28) is the relative one between two traversals of the path with the (\mathbf{q}, ω) scattering occurring at t and t' . This is averaged using the (weak) scattering probability.

Equation (28) was obtained just for the excitation of the environment caused by the electron along the path r , which will lead to its phase uncertainty $\Delta\phi_r^2$. The interaction of the path l will similarly lead to a $\Delta\phi_l^2$ and one can similarly obtain the cross terms $\langle \Delta\phi_r \Delta\phi_l \rangle = \langle \Delta\phi_l \Delta\phi_r \rangle$. The total reduction of the interference will be governed by the fluctuation of the *relative* phase

$$\langle \Delta(\phi_r - \phi_l)^2 \rangle = \langle \Delta\phi_r^2 \rangle + \langle \Delta\phi_l^2 \rangle - 2\langle \Delta\phi_l \Delta\phi_r \rangle \quad (29)$$

where in $\langle \Delta\phi_i \Delta\phi_j \rangle$ ($i, j = l, r$) the term $\exp[i\mathbf{q} \cdot (\mathbf{x}(t) -$

$\mathbf{x}(t'))]$ in equation (28) is replaced by $\exp[i\mathbf{q} \cdot (\mathbf{x}_i(t) - \mathbf{x}_j(t'))]$. The cancellation occurring among the terms in equation (29) will be seen to be of decisive importance at and below two dimensions. We will now draw a few conclusions from the above calculation:

(i) For good conductors, $\text{Im}(1/\varepsilon(q, \omega)) = \omega/4\pi\sigma$, and the probability that the state of the electron's bath was changed in the path $\mathbf{x}(t)$ is

$$\begin{aligned} P = & \frac{1}{\hbar(2\pi)^3} \int_0^{\tau_0} dt \int_0^{\tau_0} dt' \int d^3 q \int d\omega \frac{e^2 \omega}{\pi q^2 \sigma} \\ & \times \exp[i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(t')) - i\omega(t - t')] \coth \frac{\omega}{2k_B T}. \end{aligned} \quad (30)$$

For $\mathbf{x}(t) = \mathbf{x}_r(t)$, this probability (as long as $P \ll 1$) is just one-half of the uncertainty in the phase $\langle \delta\phi_r^2 \rangle$ accumulated by the right partial wave. This result is equivalent to that of Altshuler *et al*, and we shall obtain below (section 3) the phase-breaking time, τ_ϕ , from it. The present derivation demonstrates that the origin of this dephasing is in the electrostatic electron-electron interaction, and establishes the connection with the linear response of the bath.

(ii) For poor conductors $\text{Im } \varepsilon(q, \omega) \ll \text{Re } \varepsilon(q, \omega)$. Then, the q, ω integrals in equation (26) have significant contributions only from those values of q, ω in which $\text{Re } \varepsilon = 0$. A typical example is $\omega = \omega_p$, the plasma frequency.

(iii) In both cases mentioned above, the rate of dephasing crucially depends on the imaginary part of the dielectric response function. This, in turn, determines the rate at which the electron bath is excited by the interfering electron. It should be emphasized here that the polarization of the electron bath by the interfering electron, reflected in the real part of the dielectric response, does not dephase the interference. This polarization disappears when the electron leaves the polarized region and therefore it does not identify the path taken by the electron. For a general discussion of the relation between dissipation, excitations and dephasing, the reader is referred to Stern *et al* (1990a,b).

(iv) The above discussion of dephasing due to the Coulomb interaction can be easily generalized to any two-particle interaction $V(r - r')$. This is done by replacing $e^2/|r - r'|$ in equation (21) by $V(r - r')$, and following the derivations in equations (22)–(30). In particular, it is interesting to consider the case of a short-range potential, which can be approximated by $V(r - r') \propto \delta(r - r')$. For such a potential, the probability that the bath's state is changed is proportional to the density-density correlation function of the bath's electrons

$$P \propto \int_0^{\tau_0} dt \int_0^{\tau_0} dt' \langle \rho(\mathbf{x}(t), t) \rho(\mathbf{x}(t'), t') \rangle. \quad (31)$$

Thus, the intensity of the interference effects provides

here information on the density-density correlation function of the bath, which is again related to the dynamical structure factor and the dissipative part of the response.

(v) As emphasized above, the interaction of the environment with the interfering partial waves changes the state of the environment, so that it acquires information on the path taken by the interfering particle. One might then consider the case of a very slow electron traversing a piece of metal and examine the change it induces in the state of the metal. At first glance, it looks as if the state of the metal adiabatically follows the motion of the electron, so that when the electron leaves the metal, the metal is back in its initial state. However, since the excitation spectrum of the electron bath is continuous, the adiabatic argument is never applicable. It is true that the electron induces a polarization in the bath, a polarization that follows its motion adiabatically and disappears when the electron leaves the metal. But, due to the bath's continuous spectrum, this polarization has to involve an *excitation* of the bath, and this excitation does not disappear when the interaction of the bath with the interfering electron is over.

The situation is different, of course, for insulators. There, due to the gap in the excitation spectrum, a very slow electron can polarize the bath without exciting it, i.e. without identifying its path.

(vi) We have chosen to express the dephasing in terms of dynamic correlations of densities. Using the continuity equation $\rho_{q\omega} = -(q/\omega)j_{q\omega}$, one may express the latter in terms of dynamic current correlations. The longitudinal (parallel to q) components appear in the latter (due to having $\nabla \cdot j$ in the charge conservation condition which can in turn be expressed in terms of correlations of the longitudinal components of the vector potential \mathbf{A}) as in the original work of Altshuler *et al.* We believe that the presentation here makes the connection with the Coulomb-electron-electron interaction very clear.

3. Review of results in various dimensions

The final expression (30) of the previous section defines the phase uncertainty accumulated by the right partial wave. The physically meaningful object is the uncertainty in the *phase difference* between any two paths, for example, the right and left ones. It may be felt that this difference should be of the same order as each of the uncertainties $\langle \Delta\phi_1^2 \rangle$ and $\langle \Delta\phi_r^2 \rangle$. This is in fact true for $d > 2$; however, for $d \leq 2$, each of the above single-path fluctuations diverges. This unphysical divergence is cancelled by its counterpart in the mixed $\langle \Delta\phi_1 \Delta\phi_r \rangle$ terms. So here the subtraction is crucial. The divergence is a typical ‘infrared’, or low- q , effect and it is seen immediately from the $1/q^2$ in the denominator of equation (30), which at $d \leq 2$ is not cured by the phase space factor q^{d-1} from the q integration. This divergence and its remedy are in exact analogy with, for example, the by now well known anomaly in the fluctuations of the 1D and 2D lattice (for example, Imry and Gunther

1971). It is also relevant for lower-dimension superconductors.

To evaluate equation (30) and the three other terms discussed following equation (29) we start with the ω integration. The integrand will be seen in section 4 to vanish for energy transfers much larger than $k_B T$. Thus the integration produces a peak around $t' - t = 0$ whose width is $(k_B T)^{-1}$. We assume that the times of interest such as τ_ϕ and the duration of the interference experiment are much longer than $(k_B T)^{-1}$. The ω integral can then be approximated as proportional to $2\pi\delta(t - t')$. This is, in fact, the assumption that $k_B T \tau_\phi \gg 1$. This means that the width of the quasi particle excitations is much smaller than their energies, which is a basic assumption of the Fermi-liquid theory underlying much of our thinking about metals. The final results will indeed be consistent with this assumption. Summing together all the four terms of the phase uncertainty, we obtain

$$\langle \delta\phi^2 \rangle = \frac{4}{\pi^2} \int_0^{\tau_0} dt \int dk \frac{e^2 k_B T}{\sigma k^2} \sin^2[\mathbf{k} \cdot (\mathbf{x}_1(t) - \mathbf{x}_2(t))] \quad (32)$$

and τ_ϕ , the phase-breaking time, is the value of τ_0 for which the phase uncertainty is of order unity.

There are two important points that should be emphasized regarding this expression. The first is that $\langle \delta\phi^2 \rangle$ is not necessarily a linear function of time. Since the intensity of the interference term is reduced by the factor $\exp(-\frac{1}{2}\langle \delta\phi^2 \rangle)$, this means that the reduction of the interference term does not have to be a simple exponential function of time. This result is important in, for example, the analysis of the conductance of a mesoscopic ring as a function of the magnetic flux inside the ring.

The second point is the strong dimensionality dependence of the phase uncertainty. Equation (32) for $d = 1, 2$ (where d is the dimensionality of the sample) can be approximated as follows. For $\mathbf{k} \cdot (\mathbf{x}_1 - \mathbf{x}_2) \ll 1$ the dangerous k^2 denominator is compensated by the \sin^2 and this contribution is easily seen to be small. For $\mathbf{k} \cdot (\mathbf{x}_1 - \mathbf{x}_2) \gg 1$ we have an oscillatory contribution which tends to cancel out, and we remain with the average, $\frac{1}{2}$, of the sine squared which would diverge at small k , except that the integrand is cut off with $k(x_1 - x_2)$ becoming comparable with unity. Thus $k \sim 1/|\mathbf{x}_1 - \mathbf{x}_2|$ is the relevant ‘infrared’ cut-off for a divergent $\int dk k^{d-3}$. This yields, with σ multiplied by the film thickness in thin films and by the wire cross section in thin 1D wires

$$\langle \delta\phi^2 \rangle \sim \frac{e^2 k_B T}{\sigma} \int_0^{\tau_0} dt |\mathbf{x}_1(t) - \mathbf{x}_2(t)|^{2-d}. \quad (33)$$

In other words, the main contribution to the \mathbf{k} integral of equation (32) comes from $k \sim |\mathbf{x}_1(t) - \mathbf{x}_2(t)|^{-1}$, and large values of k do not contribute. Since for typical paths in a diffusive medium $|\mathbf{x}_1(t) - \mathbf{x}_2(t)| \sim \sqrt{Dt}$, we obtain for these paths

$$\langle \delta\phi^2 \rangle \sim \frac{e^2 k_B T}{\sigma} D^{(2-d)/2} t^{(4-d)/2} \quad (34)$$

and for the phase-breaking time (at which $\langle \delta\phi^2 \rangle \sim 1$)

$$\tau_\phi \sim \left(\frac{\sigma}{e^2 k_B T D^{(2-d)/2}} \right)^{2/(4-d)} \quad (35)$$

The case $d = 2$ is special in that logarithmic factors appear. A careful analysis shows that there is a logarithmic correction to τ_ϕ , but it is of the nature of $\log(\sigma d \hbar/e^2)$, d being the thickness, and no $\log T$ contribution appears.

For $d = 3$, the k integral of equation (32) diverges at the upper limit. It is cut off by the condition $Dk^2 < \omega < k_B T$, i.e. by $|k| = (k_B T/D)^{1/2}$ (see, for example, Imry *et al* 1982). Then

$$\langle \delta\phi^2 \rangle \sim \frac{e^2 k_B T}{\sigma} \left(\frac{k_B T}{D} \right)^{1/2} \tau_0 \quad (36)$$

where we assume that $(k_B T/D)^{1/2}|x_1(t) - x_2(t)| \gg 1$ for most values of t . Therefore, for $d = 3$

$$\tau_\phi \sim \frac{\sigma D^{1/2}}{e^2 (k_B T)^{3/2}}. \quad (37)$$

These results lead often, especially at low T , to stronger dephasing than those due to electron–phonon coupling. Thus, in disordered metals, interference is dephased mainly by the Coulomb interaction, or, equivalently by longitudinal fluctuations of the electromagnetic potential. Unlike the transverse fluctuations that originate in the photon modes and exist also for insulators, the longitudinal modes originate from electron–electron interactions, and are diminished when the metal becomes an insulator. The most effective fluctuations in one- and two-dimensional conductors are those of wavelengths comparable to the distance between the two interfering paths, i.e. those where $k \sim l_\phi^{-1}$. Longer wavelengths contribute to the uncertainty in each wave's phase, but keep the relative phase well defined. Shorter wavelengths make the relative phase uncertain, but their magnitude is relatively small.

Above two dimensions the phase uncertainty increases linearly with time, but in a one-dimensional system

$$\langle \delta\phi^2 \rangle \sim \left(\frac{t}{\tau_\phi} \right)^{3/2}. \quad (38)$$

with τ_ϕ given by equation (35). The 2/3 power of equation (35) has received convincing experimental confirmation by Wind *et al* (1986) and Pooke *et al* (1989). The former results for τ_ϕ^{-1} as a function of T are shown in figure 2. τ_ϕ was obtained from the weak-localization magnetoresistance.

We now discuss how the effective dimensionality of the system is determined. Consider a slab of thickness d . Intuitively, one might argue that during the time τ_ϕ the electron propagates a characteristic distance L_ϕ . For $L_\phi \ll d$ it is clear that the propagating electron will typically dephase before feeling the finite thickness, and the dephasing process will be approximately 3D-like. On the other hand, for $L_\phi \gg d$, the diffusing electron cloud fills the whole film thickness and propagates in a 2D manner before dephasing has occurred. Thus, the condition for effective low-dimensional behaviour would

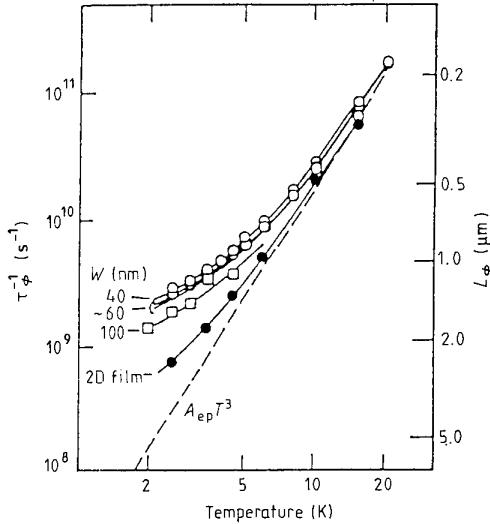


Figure 2. Phase-breaking rate versus temperature (from Wind *et al* 1986). The full curves for the wires are fits by equation (35), $d = 1$. The data for wire Ag2 ($W = 100$ nm) from 2 to 4.5 K are normalized to the R_\square and D of the Al samples to allow comparison with results for the Al wires. The full curve for the 2D Al film is a fit by the form $A'_{ee} T + A'_{ep} T^3$, with $A'_{ee} = 3.9 \times 10^8 \text{ K}^{-1} \text{ s}^{-1}$. The broken line plots the electron–phonon rate, $\propto T^3$. The scale for L_ϕ applies for the Al samples only. (Points: ○, Wires Al2a, b, c; □, wire Ag2; ●, film Al2F.)

appear to be

$$L_\phi \gg \text{appropriate length of the system.} \quad (39)$$

However, an analysis (for example that by Sivan *et al* (1994), mentioned below) of the k integration in equation (32) shows that the length $L_T = (\hbar D/k_B T)$, and not L_ϕ , is the relevant one in equation (39). The relevance of the energy ($k_B T$)-dependent length to interactions has been noted before (for example, Imry and Ovadyahu 1982, Altshuler and Aronov 1985). Since $L_\phi \gtrsim L_T$, the correct condition is more restrictive.

In a thin wire, the 1D behaviour will apply once L_T is larger than both transverse dimensions. When the wire is also of a finite length L , a further cross-over will happen once $L_T \gg L$. The system will become zero dimensional (0D), where $1/\tau_\phi$ goes like T^2 , and similar cross-overs to 0D will also happen for $d = 2, 3$ (Sivan *et al* 1994b). This result is of special interest owing to the following circumstances. Equation (39) with L_T replacing L_ϕ implies that at the cross-over to 0D, $k_B T \sim E_c$. We shall now demonstrate that the cross-over to 0D prevents the Landau Fermi-liquid theory from being violated in a narrow wire. Recall that a basic assumption of the Fermi-liquid theory is that the width of the quasiparticle excitation be much smaller than its energy. Since it is suggested physically that the above width is \hbar/τ_ϕ , this condition at temperature T is $k_B T \tau_\phi / \hbar \gg 1$. Since equation (35) implies that, at $d = 1$, $\tau_\phi \sim T^{-2/3}$, the above condition appears to be violated at low enough temperatures. Using the Einstein relation and the fact that σ in the 1D result (equation (35) with $d = 1$) is multiplied by the cross section, we first write the 1D result in the

particularly transparent fashion

$$\frac{\hbar}{\tau_\phi} \sim \left(\frac{k_B T \Delta}{\sqrt{E_c}} \right)^{2/3} \quad (40)$$

which implies that at the cross-over, $k_B T \sim E_c$, $k_B T$ is the characteristic energy of an excitation whose width satisfies $\hbar/\tau_\phi \sim E_c/g^{2/3} \ll k_B T$. Here g is the dimensionless conductance E_c/Δ . Thus at the cross-over the assumption of a good Fermi liquid ($k_B T \tau_\phi/\hbar \gg 1$) is well satisfied. Below the cross-over, the rate $1/\tau_\phi$ decreases faster than $k_B T$. Thus, the condition for a valid Fermi-liquid picture is always satisfied. This is very gratifying both on general grounds and because we have assumed it in our derivation of τ_ϕ , which is therefore self-consistent. Thus, although the 1D result *appears* to be very problematic in this respect, since \hbar/τ_ϕ decreases more slowly with temperature than the energy $k_B T$, the situation is saved by the 0D cross-over. Another way to put the above argument is by saying that the condition for breakdown of the Fermi-liquid picture is (Altshuler and Aronov 1985) $L_T > \xi$, where ξ is the localization length. The transition of 0D occurs when $L_T > L$. Thus, for a wire which is not in the localized regime (i.e. its length satisfying $L < \xi$) the former will *not* happen. The 0D cross-over occurs before $k_B T \tau_\phi/\hbar$ becomes smaller than unity, and rescues the Fermi-liquid theory. Since in the metallic limit the conductance per square of a thin film is large, the Fermi-liquid assumption is also valid in a 2D thin film.

We now discuss briefly the evaluation by Sivan *et al* (1994b) of the dephasing rate in a ‘quantum dot’ (a small finite particle), including the zero-dimensional (0D) limit. One has to calculate the relative phase fluctuation, equation (29). It has four terms, each similar to equation (28), and to evaluate them one needs averages of the type $\langle \exp[i\mathbf{q} \cdot (\mathbf{x}_i(t) - \mathbf{x}_j(t'))] \rangle$ over the diffusive motion where $i, j = r, l$ are two diffusive paths. The calculation is done by expanding an initial wavepacket which is localized at the origin, in terms of eigenfunctions of the diffusion equation with the appropriate boundary conditions (zero current through the surface) for the dot. The time dependence is then obtained by letting the wavepacket evolve with time according to the appropriate eigenvalues of the diffusion equation. It is important that the $\mathbf{q} = 0$ mode is irrelevant due to charge neutrality. The calculation is done for an electron of energy ε above the Fermi energy at $T = 0$, and as shown in the next section the effect of the Pauli principle is to limit the ω integration to the interval $[0, \varepsilon]$.

Each of the particle’s dimensions L_i defines a Thouless energy in the diffusive regime

$$E_c^i \equiv \frac{\hbar D}{L_i^2}. \quad (41)$$

For $\varepsilon \gg$ all three E_c^i , the integration is three dimensional and one obtains the 3D result, equation (37). Once $\varepsilon \lesssim$ a given E_c^i , the effective dimension is reduced in that direction(s). In the zero-dimensional limit $\varepsilon \ll E_c \equiv \min_i(E_c^i)$, one finds

$$\tau_\phi^{-1} \sim \frac{\Delta}{\hbar} \left(\frac{\varepsilon}{E_c} \right)^2. \quad (42)$$

This implies that, for energies $\varepsilon \gtrsim E_c$, the inelastic broadening (crossing over to the 3D result (37)) is enough to effectively smear the discrete spectrum ($\Delta\tau_\phi/\hbar \ll 1$). This agrees with experiments (Sivan *et al* 1994a).

4. Dephasing time versus electron-electron scattering time

Sections 2 and 3 of this paper were devoted to the calculation of the phase-breaking time, τ_ϕ , due to the electron-electron interaction, using the general principles developed in section 1. In the present section we discuss the relation between this dephasing time and the electron-electron scattering time, τ_{ee} . The latter time-scale is defined in the following manner. An electron is put in a *single-electron eigenstate* in a disordered system with a given impurity configuration. This eigenstate is characterized by an energy E . Due to the interaction of this electron with the Fermi sea, the eigenstate acquires a width. This width, averaged over all impurity configurations, is τ_{ee}^{-1} . This time-scale obviously depends on the state of the Fermi sea. The case in which the Fermi sea is at zero temperature was studied by Altshuler *et al* (1981), who obtained that $\tau_{ee}^{-1} \sim E^{d/2}$. The case of a Fermi sea at a finite temperature was discussed by Abrahams *et al* (1981) for a two-dimensional system. Other cases were studied by Schmid (1974) and by Eiler (1985).

In this section we review the results of previous calculations of τ_{ee} , with two goals in mind. The first is to relate them to the τ_ϕ calculation, and show if, when and why the two are identical. The second is to discuss an important subtlety overlooked by our calculation of τ_ϕ , namely the Pauli constraint of the energy loss of the interfering electron. Since in the approach we have taken in sections 2 and 3 the interfering electron is distinguishable from the rest of the electrons, the state to which it is scattered is not restricted to be vacant. On top of that, the amplitude for an exchange of the interfering electron with an electron from the Fermi sea is neglected. We shall try to use the τ_{ee} calculation in order to study how significant are the errors made in our present approach, and how a relatively simple recipe can semiquantitatively correct them.

We start our review of the τ_{ee} calculations with the zero-temperature case, and then extend the results to finite temperatures. The golden rule expression for the scattering rate is, in this case

$$\begin{aligned} \frac{1}{\tau_{ee}} = & \frac{2\pi}{\hbar n(0) L^d} \int_0^E d\omega \int_{-\omega}^0 d\varepsilon \sum_{\alpha\beta\gamma\delta} |V_{\alpha\beta\gamma\delta}|^2 \delta(E - \varepsilon_\alpha) \\ & \times \delta(E - \omega - \varepsilon_\beta) \delta(\varepsilon - \varepsilon_\gamma) \delta(\varepsilon + \omega - \varepsilon_\delta) \end{aligned} \quad (43)$$

where $n(0)$ is the density of states at the Fermi energy, $\alpha, \beta, \gamma, \delta$ label exact single-particle states with energies $\varepsilon_\alpha, \varepsilon_\beta, \varepsilon_\gamma, \varepsilon_\delta$ and V is the Coulomb interaction. By

diagrammatic methods Altshuler *et al* obtained the following expression for τ_{ee}^{-1}

$$\tau_{ee}^{-1} = \frac{2e^2}{\pi} \int_0^E d\omega \int \frac{d\mathbf{q}}{q^2} \frac{\sigma\omega}{\omega^2 + (Dq^2 + 4\pi\sigma)^2} \text{Re} \left(\frac{1}{i\omega + Dq^2} \right). \quad (44)$$

Before proceeding to study how this expression is related to the derivations of the previous sections, we consider its simple interpretation. The integration variable ω is the energy transfer in the scattering. It is limited by E because of the Pauli exclusion principle. We emphasize again that the electron sea is at zero temperature, so that the scattered electron cannot absorb energy from it. The integrand is composed of the imaginary part of the screened Coulomb potential

$$\frac{e^2}{q^2} \frac{\sigma\omega}{\omega^2 + (Dq^2 + 4\pi\sigma)^2} = \frac{e^2}{q^2} \text{Im} \left(\frac{1}{\varepsilon(\mathbf{q}, \omega)} \right)$$

multiplied by the ‘diffusion pole’

$$\text{Re} \left(\frac{1}{i\omega + Dq^2} \right).$$

It is known (for example, McMillan 1981) that

$$\frac{1}{\pi\hbar N(0)} \text{Re} \left(\frac{1}{i\omega + Dq^2} \right) = |\langle m | e^{i\mathbf{q} \cdot \mathbf{r}} | n \rangle|_{av}^2 \quad (45)$$

where the subscript av denotes ensemble averaging over diffusive states. Thus, equation (45) is the first-order exchange contribution to the imaginary part of the electron’s self-energy, where the perturbation is the complex potential $4\pi e^2/q^2 \varepsilon(\mathbf{q}, \omega)$. The Pauli constraints on the electron–hole excitation (for example, the limitation of its energy to be at the most ω below the Fermi energy) are all hidden in the dielectric function. Similar remarks apply to a related calculation by Giuliani and Quinn (1982) for 2D ballistic systems. According to these calculations, the ordinary $\tau_{ee}^{-1} \sim E^2$ in the pure case for $d > 2$ is replaced at $d = 2$ by $E^2 \ln |E|$.

The relation of equation (44) to the expressions discussed in the context of τ_ϕ becomes evident when we write the diffusion pole as

$$\begin{aligned} \frac{1}{i\omega + Dq^2} &= \int_0^\infty dt \exp(-Dq^2 t - i\omega t) \\ &= \tilde{N} \int_0^\infty dt \int D[\mathbf{x}(t)] \\ &\times \exp \left(- \int \frac{\dot{\mathbf{x}}^2(t') dt'}{4D} + i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(0)) - i\omega t \right) \end{aligned} \quad (46)$$

i.e. as the Laplace–Fourier transform of the average over the diffusive probability distribution of $e^{i\mathbf{q} \cdot (\mathbf{x}(t) - \mathbf{x}(0))}$, which is $\exp(-Dq^2 t)$. Here \tilde{N} is a normalization factor. The electron–electron scattering time then becomes,

defining $x(0) = 0$

$$\begin{aligned} \tau_{ee}^{-1} &= \tilde{N} \int_0^\infty dt \int D[\mathbf{x}(t)] \exp \left(- \int \frac{\dot{\mathbf{x}}^2(t') dt'}{4D} \right) \frac{2e^2}{\pi} \int_0^E d\omega \\ &\times \int \frac{d\mathbf{q}}{q^2} \text{Im} \left(\frac{1}{\varepsilon(\mathbf{q}, \omega)} \right) \text{Re} e^{i\mathbf{q} \cdot \mathbf{x}(t) - i\omega t}. \end{aligned} \quad (47)$$

In the previous section we derived an expression for the time after which an electron traversing a path $\mathbf{x}(t)$ changes the quantum state of the electron sea with which it interacts (equation (28)). Looking carefully at that expression and at equation (47), we see that τ_{ee}^{-1} is ‘almost’ the average over diffusive paths $\mathbf{x}(t)$ of equation (28) with $P(\tau_0) = 0(1)$. The main difference between the two is the limits of integration on the energy transfer variable ω . In the expression for τ_{ee} , ω is bounded by the excess energy above the Fermi level, E . In the expression for the dephasing time of the path $\mathbf{x}(t)$, it is unbounded. This difference should come as no surprise to us, in view of the ‘distinguishability’ of the interfering electron from the Fermi sea in our calculation of the dephasing time. It obviously suggests a cure to this flaw of equation (28) by limiting the ω integration from 0 to E . This demonstrates that at higher dimensions, $d > 2$, when τ_ϕ is determined by the typical time at which a single path excites the environment, it is of the same order of magnitude as τ_{ee} . At $d \leq 2$, we have the subtlety, discussed in the beginning of section 3, that the divergence in the $\langle \Delta\phi^2 \rangle$ of a single path necessitates the subtraction of the two paths, which yields the *physically meaningful* τ_ϕ .

What happens at finite temperatures? The finite-temperature expression for τ_{ee} was derived by Abrahams *et al* (1981) (note that we use a different notation). As one may expect, the sharp Pauli constraint on the energy transfer is replaced by a smooth function of ω , E and the temperature T

$$\coth \frac{\omega}{2k_B T} - \tanh \frac{\omega - E}{2k_B T}. \quad (48)$$

As hinted by the E dependence of the second term, it originates from the Pauli principle constraints of the electron of energy E . Our dephasing time calculation has yielded the $\coth(\omega/2k_B T)$ factor, but has failed to yield the $\tanh[(\omega - E)/2k_B T]$ term. Again, the reason for this failure is the neglect of the Pauli constraint on the energy loss of the interfering electron.

Our comparison of the τ_ϕ and τ_{ee} calculation brings us then to the following recipe for the calculation of the diffusive electron–electron scattering time τ_{ee} when it is meaningful, i.e. for $d > 2$: first, calculate the time it takes an electron whose path is $\mathbf{x}(t)$ to change the state of the Fermi sea it couples to. Second, average over all diffusive paths $\mathbf{x}(t)$. Third, correct for the Pauli constraints by using the thermal factor equation (48). At $d \leq 2$ only the time for relative dephasing, τ_ϕ , is physically meaningful.

Being aware now of the approximations that were made in our calculation of τ_ϕ , we may examine what are the conditions under which *this* calculation is valid, i.e.

that our neglect of Pauli's principle does not significantly affect the final result. The answer to this question is easily found by an examination of equation (48). The second term of the thermal factor is negligible when $\omega \ll k_B T$. Thus, if the dephasing is dominated by energy transfer ω which is much smaller than the temperature, then the neglect of Pauli's principle is not significant. As seen in section 3, this is the case for low dimensions, $d = 1, 2$. At low dimensions the time it takes each of the paths $x_1(t), x_2(t)$ to change the quantum state of the Fermi sea diverges due to small momentum and energy transfers. It is only the strong overlap of the excitations induced by each of the two paths that makes the dephasing time finite. Thus, the Fermi constraints of the interfering electron are unimportant. For $d = 3$ the situation is different. The dephasing time is dominated by energy transfers of the order of the temperature. Now, if the energy $E > k_B T$, again the $\tanh[(\omega - E)/2k_B T]$ term does not significantly affect the final result. It is only for $E \ll k_B T$ that our calculation of the dephasing time becomes wrong. Since the interfering electrons we discuss typically have an energy $k_B T$ above the Fermi energy, our calculation of the dephasing time still remains qualitatively valid.

Acknowledgments

We thank Y Aharonov, A G Aronov and U Sivan for collaboration on some of the work reported here, and T D Schultz, B L Altshuler, D E Khmel'nitskii, U Smilansky and A. Yacoby for essential discussions. Research at the Weizmann Institute of Science was supported by the Minerva Foundation and the German-

Israeli Foundation (GIF). AS is grateful to the Harvard Society of Fellows for financial support.

References

- Abrahams E, Anderson P W, Lee P A and Ramakrishnan T V 1981 *Phys. Rev. B* **24** 6783
- Altshuler B L and Aronov A G 1985 *Electron-Electron Interactions in Disordered Systems* ed A L Efros and M Pollak (Amsterdam: North-Holland) p 1
- Altshuler B L, Aronov A G and Khmel'nitskii D E 1981 *Solid State Commun.* **39** 619
- 1982 *J. Phys. C: Solid State Phys.* **15** 7367
- Caldeira A O and Leggett A J 1983 *Ann. Phys., NY* **149** 374
- Eiler W 1985 *Solid State Commun.* **56** 11
- Feynman R P and Vernon F L 1963 *Ann. Phys., NY* **24** 118
- Giuliani G F and Quinn J J 1982 *Phys. Rev. B* **26** 4421
- Holstein T 1961 *Phys. Rev.* **124** 1329
- Imry Y, Gefen Y and Bergmann D J 1982 *Phys. Rev. Lett. B* **26** 3436
- Imry Y and Gunther L 1971 *Phys. Rev. B* **3** 3939
- Imry Y and Ovadyahu Z 1982 *Phys. Rev. Lett.* **49** 841
- McMillan W L 1981 *Phys. Rev. B* **24** 2739
- Nozières P 1964 *Interacting Fermi Systems* (New York: Benjamin)
- Pooke L, Paquin N, Pepper M and Gundlach A J 1989 *J. Phys.: Condens. Matter* **1** 3289
- Schmid A 1974 *Z. Phys.* **271** 251
- Sivan U, Aronov A G and Imry Y 1994b *Europhys. Lett.* submitted
- Sivan U, Milliken F P, Milkove K, Rishton S, Lee Y, Hong J M, Boegli V, Kern D and de Franza M 1994a *Europhys. Lett.* **25** 605
- Stern A, Aharonov Y and Imry Y 1990a *Phys. Rev. A* **40** 3436
- 1990b *Quantum Coherence in Mesoscopic Systems* ed B Kramer (New York: Plenum) p 99
- Wind S, Rooks M J, Chandrasekhar V and Prober D E 1986 *Phys. Rev. Lett.* **57** 633

Transport spectroscopy on a single quantum dot

J Weis, R J Haug, K von Klitzing and K Ploog†

Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart,
Germany

Abstract. Lateral transport through a quantum dot defined by the split-gate technique in a two-dimensional electron gas is investigated as a function of backgate voltage *and* emitter–collector bias voltage. This measurement technique allows us to identify the regimes of single-electron tunnelling. Within these regimes, excited states of the electron system in the quantum dot provide additional transport channels which can be classified as being opened in resonance with the Fermi level of either the emitter or the collector. The method of transport spectroscopy is discussed. When performing spectroscopy in a magnetic field, one has to take into account that the magnetic field affects not only the electronic states of the quantum dot but also the electronic states in the electrodes surrounding the quantum dot.

1. Introduction

In a quantum dot, electrons are confined in all three dimensions, leading to a discrete energy spectrum of the electronic states. Coupling the quantum dot to two electron reservoirs by weak tunnel junctions allows the investigation of electron transport through the quantum dot. At low bias voltage between the emitter and the collector reservoir, and at low temperature, transport through the quantum dot is normally suppressed due to the lack of the energy required to add an electron to the quasi-isolated quantum dot. This energy is mainly given by the charging energy of the capacitor formed by the quantum dot and the electrodes. This effect is known as the Coulomb blockade of tunnelling [1]. By changing the electrostatic potential of a gate electrode, capacitively coupled to the quantum dot, this energy barrier can be overcome, allowing transport through the quantum dot. Within this transport regime, the number of electrons in the quantum dot can change only by one at a time. This is called the regime of single-electron tunnelling. By changing the gate voltage further, the transport is blocked again, having increased the number of electrons in the quantum dot by one. In this arrangement of a single-electron-tunnelling transistor, transport spectroscopy through the quantum dot is done by varying both the gate voltage *and* the emitter–collector bias. Excited states of the quantum dot become accessible, providing additional transport channels [2–4].

† Present address: Paul-Drude Institut für Festkörperferelektronik, Hausvogteiplatz 5-7, 10117 Berlin, Germany.

2. The model

The excess charge Q_i on a metallic electrode i with the electrostatic potential V_i depends linearly on the differences of the electrostatic potentials $V_i - V_j$ for all electrodes j : $Q_i = \sum_{j \neq i} C_{ij}(V_i - V_j)$. The sum is to be taken over all electrodes. The capacitance coefficients C_{ij} are given by the geometrical arrangement and the dielectric material outside the metallic electrodes. Having an excess charge Q_0 on the island electrode and the electrostatic potentials of all other electrodes being fixed, the electrostatic potential of the island electrode is given by

$$V_0 = \sum_{j \neq 0} \frac{C_{0j}}{C_{0\Sigma}} V_j + \frac{Q_0}{C_{0\Sigma}}. \quad (1)$$

$C_{0\Sigma} = \sum_{j \neq 0} C_{0j}$ is the total capacitance of the island electrode. The excess charge on the *isolated* electrode can change only by an integer multiple of the electron charge $-e$ due to the quantization of charge. Thus, taking ΔN electrons from the reference electrode with electrostatic potential equal to zero and adding these ΔN electrons to the initially electrically neutral island, the electrostatic energy of the electrode arrangement changes by $E_{el} = \int_0^{-\Delta Ne} V_0(Q_0) dQ_0$ [5], which leads with equation (1) to

$$E_{el}(\Delta N; \{V_j\}) = -\Delta Ne \sum_{j \neq 0} \frac{C_{0j}}{C_{0\Sigma}} V_j + \frac{(\Delta Ne)^2}{2C_{0\Sigma}}. \quad (2)$$

Here, the electrostatic potentials of all electrodes except that of the island electrode are kept fixed. E_{el} depends on *all* electrostatic potentials and this is denoted by $\{V_j\}$. Due to equation (2) the energy necessary to put ΔN

electrons on the initially electrically neutral island is given by two terms. The first term describes the energy gained for ΔN electrons being moved from the reference electrode to the island in the electrostatic field caused by the electrodes. The second term takes into account the work necessary to separate the negative charge $-\Delta Ne$ from its positive image charge, which is spread over the surface of the other electrodes. It describes the charging energy of the capacitor that is formed by the island electrode and all other electrodes.

If the island is a quantum dot, the confinement and therefore the discrete energy spectrum and the Pauli principle have to be taken into account. In a first approach, a discrete one-particle energy spectrum has been added to the electrostatic energy of equation (2) (see [6]). In a more general form, the Hamiltonian of N electrons enclosed by an electrostatic potential, which is given by a fixed charge distribution $\rho_{\text{ion}}(\mathbf{r})$ and an arrangement of M electrodes, can be expressed by the electrostatic Green function $G(\mathbf{r}, \mathbf{r}')$:

$$H(N) = \sum_{s=1}^N \frac{\mathbf{p}_s^2}{2m} - e \sum_{s=1}^N \Phi_{\text{ext}}(\mathbf{r}_s) + \frac{1}{2} \sum_{s=1}^N \sum_{\substack{s'=1 \\ s' \neq s}}^N e^2 G(\mathbf{r}_s, \mathbf{r}_{s'}). \quad (3)$$

The Green function $G(\mathbf{r}, \mathbf{r}')$ is defined as the solution of

$$\begin{aligned} \nabla_{\mathbf{r}} [\epsilon_0 \epsilon(\mathbf{r}) \nabla_{\mathbf{r}'} G(\mathbf{r}, \mathbf{r}')] &= -\delta(\mathbf{r} - \mathbf{r}') & (\mathbf{r} \in V) \\ G(\mathbf{r}, \mathbf{r}') &= 0 & (\mathbf{r} \in S) \end{aligned} \quad (4)$$

where V describes the volume between the electrodes, S the surface of all electrodes and $\epsilon(\mathbf{r})$ the dielectric medium in V .

The electrostatic potential $\Phi_{\text{ext}} = \Phi_{\text{ion}} + \Phi_{\text{gates}} + \Phi_{\text{im}}$ is independent of the number of electrons in the quantum dot:

$$\Phi_{\text{ion}}(\mathbf{r}) = \int_{V'} \rho_{\text{ion}}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') d^3 r' \quad (5)$$

$$\Phi_{\text{gates}}(\mathbf{r}) = \sum_{j=1}^M \int_{S'_j} [\epsilon_0 \epsilon(\mathbf{r}') \nabla_{\mathbf{r}'} G(\mathbf{r}', \mathbf{r})] dS'_j / V_j \quad (6)$$

$$\Phi_{\text{im}}(\mathbf{r}) = \frac{e^2}{2} G'(\mathbf{r}, \mathbf{r}).$$

$\Phi_{\text{im}}(\mathbf{r})$ describes the potential seen by the electron at \mathbf{r} caused by its own image charges. $G'(\mathbf{r}, \mathbf{r}')$ is given by $G(\mathbf{r}, \mathbf{r}')$ excluding the divergent self-energy part [7].

By solving the many-particle Schrödinger equation $H(N)|N, l\rangle = E(N, l)|N, l\rangle$ one obtains the energy spectrum $E(N, l)$. Here, the index l stands for a set of quantum numbers that characterizes the N -electron states. The index l is used to number the states $|N, l\rangle$, starting from $l = 0$ for the ground state. Due to the electrodes neighbouring the quantum dot, an effective electron-electron interaction occurs in equation (3), taking into account screening effects, if the intradot distances are comparable to the distances between the quantum dot and the electrodes. This term is analogous to the last term in equation (2), and the term given in equation (6) is analogous to the first term in equation (2).

Allowing an electron exchange between the quantum dot and a weakly coupled electron reservoir with the electrochemical potential μ_E , the quantum dot can reach an equilibrium with the reservoir. Then, the probability $P(N, l)$ of finding the electron system in the quantum dot in the state $|N, l\rangle$ is given by the Gibbs distribution function [5]. The probability $P(N)$ for N electrons in the quantum dot is the probability of finding the quantum dot in any state of the N -electron system,

$$\begin{aligned} P(N) = \sum_{l=0}^{\infty} P(N, l) &= Z^{-1} \exp\left(-\frac{E(N, 0) - N\mu_E}{k_B T}\right) \\ &\times \sum_{l=0}^{\infty} \exp\left(-\frac{E(N, l) - E(N, 0)}{k_B T}\right) \end{aligned} \quad (7)$$

where Z is the partition function

$$Z = \sum_{\text{all } |N, l\rangle} \exp\left(-\frac{E(N, l) - N\mu_E}{k_B T}\right). \quad (8)$$

Because of the first exponential factor in equation (7), thermal fluctuations of the number of electrons in the quantum dot are suppressed at low temperature, i.e. it is fixed to N , if

$$\mu(N+1; \{V_i\}) - \mu_E \gg k_B T \quad (9)$$

$$\mu(N; \{V_i\}) - \mu_E \ll k_B T. \quad (10)$$

Here, $\mu(N; \{V_i\}) \equiv E(N, 0; \{V_i\}) - E(N-1, 0; \{V_i\})$ is the difference in the ground-state energies of the N - and the $(N-1)$ -electron systems at fixed electrostatic potentials $\{V_i\}$. The energy barrier for changing the number of electrons in the quantum dot is maximal if μ_E lies in the middle between $\mu(N+1; \{V_i\})$ and $\mu(N; \{V_i\})$. For metallic systems, this maximum energy barrier is $e^2/2C_0\Sigma$. At certain values of $\{V_i\}$ where $\mu(N; \{V_i\}) = \mu_E$, the number of electrons in the quantum dot is fluctuating between N and $N-1$ even at zero temperature.

For transport investigations, the quantum dot has to be coupled to a second reservoir with an electrochemical potential μ_C that is below μ_E . If the quantum dot is in the ground state of the N -electron system, transport through the quantum dot is blocked if the minimal energy $\mu(N+1; \{V_i\}) - \mu_E$, necessary to add the $(N+1)$ th electron to the quantum dot, and the minimal energy $\mu_C - \mu(N; \{V_i\})$, necessary to remove the N th electron, greatly exceed $k_B T$. Thus, an electron transport process from emitter to collector, changing the number of electrons in the quantum dot in the meantime, is suppressed if the relation

$$\mu(N+1; \{V_i\}) > \mu_E > \mu_C > \mu(N; \{V_i\}) \quad (11)$$

holds [6]. Condition (11) is shown graphically in figure 1(α). (This Coulomb blockade of tunnelling does not exist even at zero temperature for multi-electron tunnelling processes, which change the electron number in the quantum dot only virtually [8].) By changing

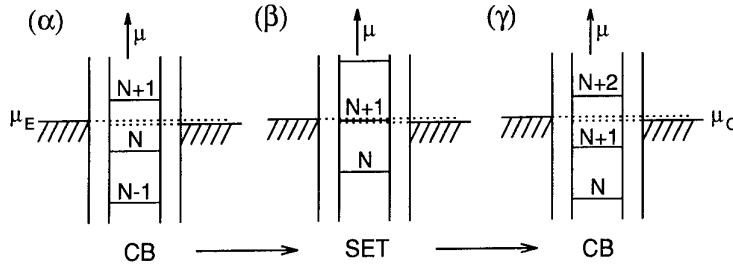


Figure 1. Energetic shift of $\mu(N; \{V_i\})$ relative to μ_E and μ_C by increasing a gate voltage V_j . Coulomb blockade (CB) occurs in cases (α) and (γ), and single-electron tunnelling (SET) takes place in case (β).

a gate voltage, the energy barrier is overcome and electron transport through the quantum dot occurs if $\mu(N + 1; \{V_i\})$ lies between μ_E and μ_C [6]. Single-electron tunnelling occurs if

$$\mu(N + 2) > \mu_E \geq \mu(N + 1) \geq \mu_C > \mu(N) \quad (12)$$

(shown graphically in figure 1(β)). On increasing the gate voltage further, the number of electrons becomes fixed at $(N + 1)$ (see figure 1(γ)). Therefore, when increasing the gate voltage at vanishingly small bias between μ_E and μ_C , a series of conductance peaks are observable and these are called Coulomb blockade oscillations. The resonance condition for a conductance peak is given by $\mu_E \approx \mu(N; V_j) \approx \mu_C$. The resonance condition for the next conductance peak is given by $\mu_E \approx \mu(N + 1; V_j + \Delta V_j) \approx \mu_C$. For small changes of the electrostatic potential V_j the energy $E(N, l)$ is shifted linearly with V_j , which can be described by $\partial E(N, l)/\partial V_j = -\alpha_j(N, l)Ne$. For a metallic system, $\alpha_j(N, l)$ is given by the capacitance ratio $C_{0j}/C_{0\Sigma}$. Assuming that $\mu(N; V_j + \Delta V_j) = \mu(N; V_j) - (C_{0j}/C_{0\Sigma})e\Delta V_j$, the distance ΔV_j between adjacent conductance peaks is

$$\frac{C_{0j}}{C_{0\Sigma}}e\Delta V_j = \mu(N + 1; \{V_i\}) - \mu(N; \{V_i\}). \quad (13)$$

For a metallic system, the distance is periodic: $\Delta V_j = e/C_{0j}$. For a quantum dot system, deviations from this periodicity are expected and are clearly observable in magnetic-field-dependent measurements [9].

3. The sample

To define the quantum dot, metallic split gates, shown schematically in the inset of figure 2(b), were deposited on top of a Hall bar etched in a GaAs/Al_{0.33}Ga_{0.67}As heterostructure with a 2DEG (electron density: $3.4 \times 10^{15} \text{ m}^{-2}$, mobility: $60 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ at a temperature of 4.2 K). The diameter of the area between the tips of the gates is about 350 nm. In addition to these top gates, a metallic electrode (back-gate) on the reverse side of the undoped substrate was used to change the electrostatic potential of the quantum dot. A schematic

cross section is shown in figure 2(a). The distance between the 2DEG and the top gates was 86 nm; that between the 2DEG and the back gate was 0.5 mm. The sample was mounted in a ³He/⁴He dilution refrigerator with a base temperature of 22 mK. The two-terminal conductance through the quantum dot was measured by using an AC lock-in technique at a frequency of 13 Hz and an effective AC source-drain voltage of 5 μ V. In addition to this latter voltage, a DC voltage V_{DS} in the range of mV could be applied. For $V_{DS} \neq 0$ mV, the differential conductance is measured by the lock-in technique. For the measurements, the top gates were kept at fixed voltages (around -0.7 V). The tunnelling barriers could be tuned by slight changes in the voltages applied to the different top-gate fingers. Here, data are presented where the conductance peaks, shown in figure 2(b), were tuned to maximum amplitude by tuning the tunnelling barriers.

4. Transport spectroscopy

A typical curve of conductance versus back-gate voltage is shown in figure 2(b). Only a few such well separated conductance resonances are observable in our system. For more negative back-gate voltages, the conductance is completely suppressed, whereas for more positive back-gate voltages the widths of the conductance resonances are broadened and a finite conductance is measured between adjacent peaks. These series of conductance peaks are interpreted in the following as Coulomb blockade oscillations. The energetic situations, shown schematically in figure 1, are found in figure 2(b) marked by (α), (β) and (γ).

In figure 3(a) the differential conductance dI/dV_{DS} is shown as a function of the back-gate voltage V_{BS} for the different bias voltages V_{DS} (between -3 mV and 3 mV in 0.1 mV steps). In the linear grey-scale plot, white regions correspond to dI/dV_{DS} below $-0.1 \mu\text{S}$ and black ones to dI/dV_{DS} above $2 \mu\text{S}$. For clarity the main structures visible in figure 3(a) are sketched in figure 3(b). At vanishingly small V_{DS} the conductance resonances are observed. By increasing the absolute value of V_{DS} , the range in back-gate voltage V_{BS} where transport through the quantum dot occurs

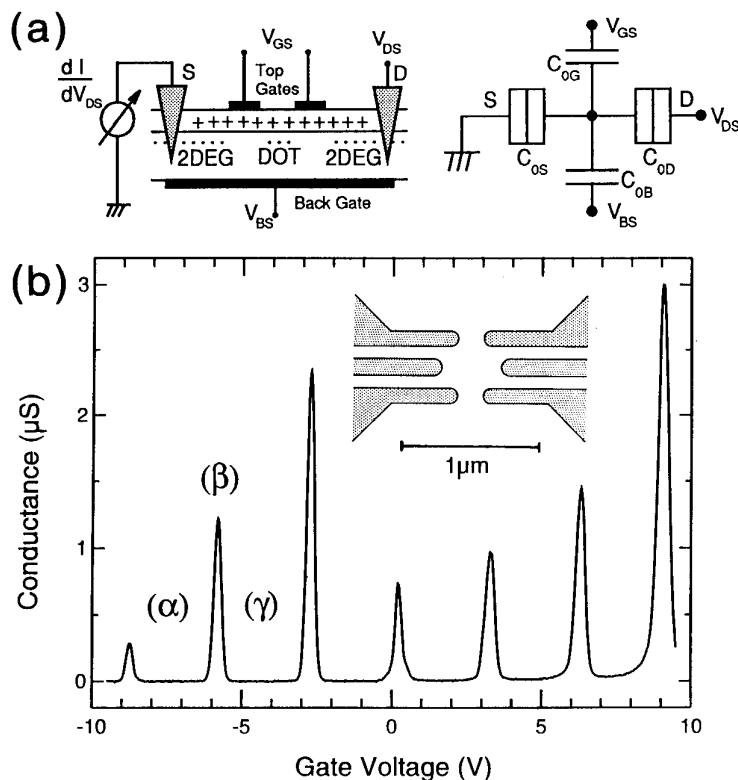


Figure 2. (a) Left: schematic cross section of the sample with applied voltages. Right: capacitor circuit of the quantum dot. (b) Conductance versus back-gate voltage. Inset: schematic diagram of the centre of the split-gate structure used to define the quantum dot.

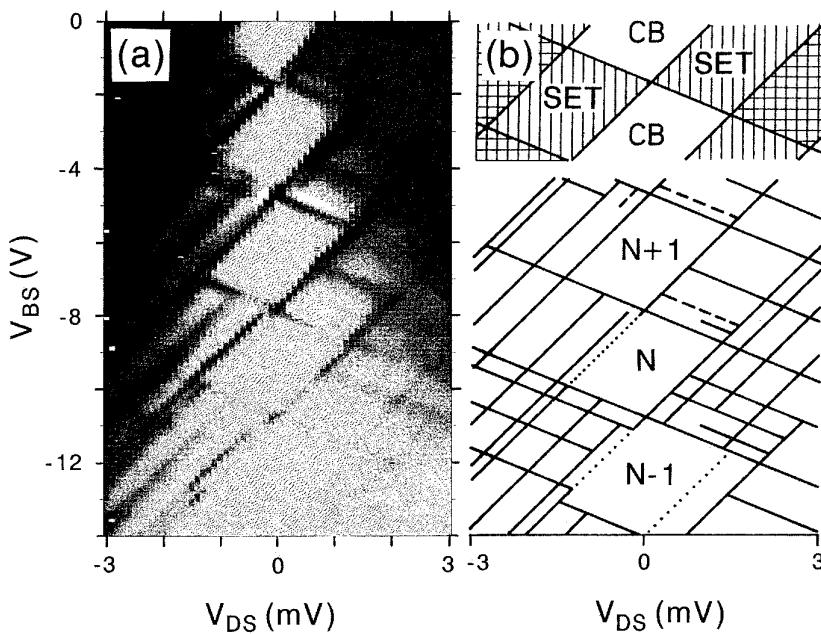


Figure 3. (a) Differential conductance dI/dV_{DS} given on a linear grey scale (white $\leq -0.1 \mu\text{S}$, black $\geq 2 \mu\text{S}$) as a function of back-gate voltage V_{BS} for different bias voltages V_{DS} . (b) Top: regions of SET are hatched; regions where the number of electrons can change by two at a time are cross-hatched. Lower part: the main structures visible in (a) are sketched. Broken lines show the regime of negative differential conductance; dotted lines show suppressed conductance.

is broadened linearly with $|V_{DS}|$. These regions of transport enclose almost rhombically shaped regions between them, where transport through the quantum dot is blocked (Coulomb blockade (CB) regime). The boundaries between transport and blockade regimes in the V_{BS} versus V_{DS} plane are defined by the conditions $\mu_E = \mu(N+1; V_{BS}^l, V_{DS})$ and $\mu(N+1; V_{BS}^u, V_{DS}) = \mu_C$ respectively, where $\mu_E - \mu_C = e|V_{DS}|$. As seen in figure 3 these resonance positions V_{BS}^u and V_{BS}^l shift linearly when changing V_{DS} . This implies that the shifts of the electrostatic potential of the quantum dot can be modelled by a capacitance circuit (shown in figure 2(a)). Therefore, the two slopes dV_{BS}/dV_{DS} , characterizing the boundaries between transport and blockade, are given by $-C_{0D}/C_{0B}$ and $(C_{0\Sigma} - C_{0D})/C_{0B}$ respectively. From both slopes, the scaling factor $\alpha = C_{0B}/C_{0\Sigma}$ between the change $\Delta\Phi$ of electrostatic potential of the quantum dot and the change ΔV_{BS} of back-gate voltage V_{BS} is obtained ($\alpha = \Delta\Phi/\Delta V_{BS} = (4.5 \pm 0.2) \times 10^{-4}$). This implies, using equation (13), that the differences $\mu(N+1; \{V_i\}) - \mu(N; \{V_i\})$ in our dot are about 1.3 meV.

The regimes of single-electron tunnelling (SET), defined by equation (12) and shown schematically in the top part of figure 3(b), can be identified in the V_{DS} versus V_{GS} plane by extrapolating the boundaries between transport and blockade regimes. Within the single-electron tunnelling regimes, the grid-like structure in figure 3 is observable, indicating additional transport channels through the quantum dot at finite bias voltage V_{DS} . What is the reason for these channels? During transport at vanishingly small V_{DS} , the quantum dot changes between the ground states of two different electron systems (e.g. between an N - and an $(N+1)$ -electron system). At finite bias voltage, excited states for both electron systems become accessible, providing new tunnelling channels through the quantum dot [2–4]. There are two possibilities: a new channel opens at the emitter side or a new channel opens at the collector side. The condition for opening a channel in resonance to the Fermi level of the *emitter* is

$$\begin{aligned}\mu_E &= E(N+1, k; \{V_i\}) - E(N, l; \{V_i\}) \\ &> \mu(N+1; \{V_i\}) \geq \mu_C.\end{aligned}\quad (14)$$

Here, the $(N+1)$ th electron is added to the N -electron state $|N, l\rangle$ in the quantum dot, leading to an excited state $|N+1, k\rangle$ of the $(N+1)$ -electron system. The condition for opening a channel in resonance to the Fermi level of the *collector* is

$$\begin{aligned}\mu_C &= E(N+1, k'; \{V_i\}) - E(N, l'; \{V_i\}) \\ &< \mu(N+1; \{V_i\}) \leq \mu_E.\end{aligned}\quad (15)$$

It becomes possible that the quantum dot is left in an excited state $|N, l'\rangle$ of the N -electron system, as the $(N+1)$ th electron leaves the quantum dot, which has been in the state $|N+1, k'\rangle$. The conditions (14) and (15) can be distinguished as the resonance position V_{BS} for the opening of a new channel from the emitter (to the collector) shifts to negative (positive) V_{BS} with

increasing $|V_{DS}|$, i.e. parallel to the boundary V_{BS}^l (V_{BS}^u) of the SET regime.

Equation (14) describes the situation when an excited state of the $(N+1)$ -electron system is reached, whereas equation (15) describes the condition for an excited state of the N -electron system becoming accessible. This implies that spectroscopy of the energy differences of an electron system with the same number of electrons is possible. This is only true if a fast and complete relaxation to the ground state of the confined electron system in the quantum dot occurs before the next tunnelling process through one of the barriers starts. Then in equation (14), the initial state $|N, l\rangle$ for adding an electron is always the ground state $|N, 0\rangle$, whereas in equation (15), the initial state $|N+1, k'\rangle$ for removing an electron is always the ground state $|N+1, 0\rangle$. Therefore, in the case of a fast and complete relaxation process, the two systems (N or $N+1$ electrons) can be distinguished, allowing the spectroscopy of the $(N+1)$ -electron system (condition (14)) and of the N -electron system (condition (15)) separately within one SET regime. But features in figure 3, which are discussed in [11], indicate that at least for some excited states relaxation does not occur within the mean time $\Delta t = e/I$ between successive electrons passing through the quantum dot. So, what can be obtained from the differences of the V_{BS} positions of differential conductance peaks shifting parallel in their V_{BS} positions with increasing $|V_{DS}|$?

To answer this, let us discuss spectroscopy at fixed bias voltage V_{DS} for the general case by using the resonance to the Fermi level of the emitter (described by (14)). Having N electrons enclosed in the quantum dot, the lower boundary to the transport regime is reached at V_{BS}^l by increasing V_{BS} if

$$\mu(N+1; V_{BS}^l) = E(N+1, 0; V_{BS}^l) - E(N, 0; V_{BS}^l) = \mu_E.$$

On increasing the gate voltage V_{BS} further by ΔV_{BS} , within the SET regime a new transport channel is opened from the emitter side if

$$E(N+1, k; V_{BS}^l + \Delta V_{BS}) - E(N, l; V_{BS}^l + \Delta V_{BS}) = \mu_E.$$

With $\partial E(N, l)/\partial V_{BS} = -(C_{0B}/C_{0\Sigma})Ne$, the difference ΔV_{BS} for opening a new transport channel is given by

$$\begin{aligned}e \frac{C_B}{C_\Sigma} \Delta V_{BS} &= E(N+1, k) - E(N, l) \\ &\quad - [E(N+1, 0) - E(N, 0)].\end{aligned}\quad (16)$$

Therefore, the interpretation of the structure within a SET regime is complex. In particular, further investigations are necessary to obtain information about the selection rules beyond energy conservation which allow the transitions between an N - and an $(N+1)$ -electron state by adding an electron to the quantum dot, e.g. the influence of the spin quantum number [12].

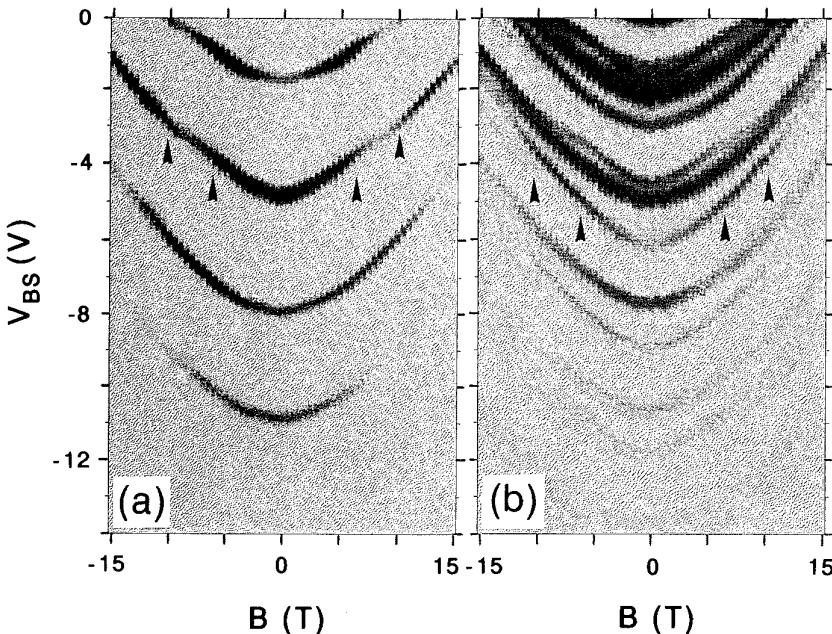


Figure 4. (a) Differential conductance dI/dV_{DS} given on a linear grey scale (white $\leq -0.1 \mu\text{S}$, black $\geq 2 \mu\text{S}$) as a function of back-gate voltage V_{BS} for different magnetic fields B . The bias voltage V_{DS} was 0 mV (see figure 2). (b) Similar measurement to that in (a), but at a bias voltage $V_{DS} = -0.7 \text{ mV}$.

5. Transport spectroscopy with a magnetic field

Applying a magnetic field to the quantum dot system changes the energy spectrum of the quantum dot. The change in $E(N+1, k) - E(N, l)$, induced by the magnetic field B , can be compensated by a change in the applied gate voltage V_{BS} . Therefore, the values of the applied gate voltage V_{BS} , which fulfil the resonance condition (14) or (15) for opening transport channels, are shifted. McEuen *et al* [9] performed this kind of transport spectroscopy in a magnetic field orientated perpendicular to the plane of the disc-like quantum dot created in an AlGaAs/GaAs heterostructure. A similar spectrum was obtained by Ashoori *et al* [10] using a capacitance measurement technique. Because of the weak confinement in the plane perpendicular to the magnetic field orientation, the spectrum is dominated by orbital effects. On the contrary, measurements here were performed in a magnetic field B orientated parallel to the plane of the 2DEG, parallel to the current direction. Because of the strong confinement in the growth direction of the heterostructure, orbital effects are diminished and the observed magnetic field dispersions are less complex.

In figure 4(a) the differential conductance dI/dV_{DS} measured at $V_{DS} = 0 \text{ mV}$ is plotted on a linear grey scale (white regions correspond to $dI/dV_{DS} \leq 0 \mu\text{S}$, black ones to $dI/dV_{DS} \geq 2 \mu\text{S}$) as a function of the back-gate voltage V_{BS} for different values of the magnetic field B between -15 T and 15 T in 0.5 T steps. In figure 4(b) similar measurements are shown for $V_{DS} = -0.7 \text{ mV}$. In this figure the gate-voltage regions, where

single-electron tunnelling occurs, are broadened since the difference between the electrochemical potentials of emitter and collector has been increased. Additional transport channels due to excited states of the quantum dot become accessible and the magnetic field dispersions of V_{BS} for opening these channels are visible. The magnetic field behaviour is symmetric on changing the sign of the magnetic field B . The upper and lower boundaries $V_{BS}^l(B)$ and $V_{BS}^u(B)$ of each SET regime show the same magnetic field dispersion as predicted by the respective conditions $\mu(n; V_{BS}, V_{DS}, B) = \mu_E$ and $\mu(n; V_{BS}^u, V_{DS}, B) = \mu_C$, where the difference $\mu_E - \mu_C = e|V_{DS}|$ is independent of the magnetic field. This is the same magnetic field dispersion that is observed at vanishingly small bias voltage $|V_{DS}|$ in figure 4. The kinks observable in figure 4 can be identified with crossings of different magnetic field dispersions in figure 4(b) (marked by arrows). Therefore, kinks in the magnetic field dispersion of the boundaries V_{BS}^u and V_{BS}^l of a SET regime indicate a change in the character of the ground state of the N - or $(N+1)$ -electron system (a change in the quantum numbers).

Different magnetic field dispersions are visible in figure 4(b). In principle, further information about the magnetic field dependence of the energy differences $E(N+1, k) - E(N, l)$ between states of N - and $(N+1)$ -electron systems of the quantum dot should be obtained. But in interpreting the shift of V_{BS} , which fulfils the condition (14) or (15) one has to be careful. On the one hand the magnetic field changes the states of the quantum dot, but on the other hand it also affects the states of the 2DEG electrodes, acting as leads, and of the metal electrodes, used as gates. The electronic

states of the electrodes are energetically shifted and the density of states is changed by the magnetic field. The differences of the *electrostatic* potentials of the electrodes are *not* given by the external electrical circuit, but the differences of the *electrochemical* potentials of the electrodes are given by the *external applied* voltages (V_{BS} , V_{DS} etc). The electrical circuit serves as a reservoir for each electrode, and therefore the electrochemical potential of each electrode is fixed. By changing the magnetic field, electrons are exchanged between the electrodes and their reservoirs to reach an equilibrium. Because of this electron exchange, an intrinsic voltage $V_i^c(B)$ is built up between the electrodes and their reservoirs, leading to a shift of the electrostatic potential of the quantum dot by $(C_i/C_\Sigma)\Delta V_i^c(B)$. The shift $(\partial V_{BS}/\partial B)\delta B$ of the applied gate voltage V_{BS} required to remain at the resonance position can be obtained from a differentiation of equations(14) and (15) respectively:

$$e \frac{C_G}{C_\Sigma} \frac{\partial V_{BS}}{\partial B} = \frac{\partial}{\partial B} (E(N+1, k) - E(N, l))|_{V_i} - e \sum_k \frac{C_k}{C_\Sigma} \frac{\partial V_k^c}{\partial B}. \quad (17)$$

Equation (17) takes into account not only the *explicit* magnetic field dependence of the electronic states in the quantum dot, but also the *implicit* magnetic field dependence due to the change of the electrostatic potential of the electrodes. This is the essential result of this consideration: by measuring the magnetic field dispersion of the applied gate voltage $V_{BS}(B)$, fulfilling the resonance condition for opening a new transport channel, not only is the change in $E(N+1, k) - E(N, l)$ at fixed $\{V_i\}$ measured, but also the change in the contact voltages $\{V_i^c\}$ induced by the magnetic field has to be taken into account. Therefore, if $\partial V_i^c(B)/\partial B$ is unknown, the interpretation of the magnetic field dispersion is difficult. As a consequence of $\mu_i = -eV_i + \mu_i^{ch} = \text{const}$, the change of the electrostatic potential V_i is given by the change of the chemical potential μ_i^{ch} of the electrode i , i.e. $e\partial V_i^c/\partial B = \mu_i^{ch}/\partial B$. The last term in equation (17) gives a contribution to all magnetic field dispersions. Consequently, direct information about the magnetic field dispersion of $E(N+1, k) - E(N, l)$ at fixed electrostatic potentials $\{V_i\}$ can be obtained from a direct comparison of two magnetic field dispersions.

6. Conclusion

Transport spectroscopy through a quantum dot has been discussed. At finite bias voltage, transport channels, using excited states of the electron system in the quantum dot, can be classified as being opened in resonance to the Fermi level of the emitter or of the collector.

In the first case an excited state of the $(N+1)$ -electron system is reached, whereas in the second case an excited state of the N -electron system is reached. Because of the absence of complete relaxation processes, the spectroscopy of the confined electron system in the quantum dot is complicated, since the tunnelling processes do not always start from a ground state in the quantum dot. When performing transport spectroscopy in a magnetic field, one has to take into account that the electrochemical potential differences and not the electrostatic potential differences are fixed by the electrical circuit.

Acknowledgments

We gratefully acknowledge stimulating discussions with D Pfannkuche, R Blick, H Pothier, D Weinmann, W Häusler, and P Maksym. We thank M Riek, A Gollhardt and F Schartner for their expert help with the sample preparation, and G Ernst and Y Kershaw for reading the manuscript. Part of this work has been supported by the Bundesministerium für Forschung und Technologie.

References

- [1] Special Issue on 'Single Charge Tunnelling' 1991 *Z. Phys. B* **85** (3)
Grabert H and Devoret M H (ed) 1992 *Single Charge Tunneling (NATO ASI Series B Vol 294)* (New York: Plenum)
- [2] Johnson A T, Kouwenhoven L P, de Jong W, van der Vaart N C, Harmans C J P M and Foxon C T 1992 *Phys. Rev. Lett.* **69** 1592
- [3] Weis J, Haug R J, von Klitzing K and Ploog K 1992 *Phys. Rev. B* **46** 12837
- [4] Foxman E B, McEuen P L, Meirav U, Wingreen N S, Meir Y, Belk P A, Belk N R, Kastner M A and Wind S J 1993 *Phys. Rev.* **47** 10020
- [5] Beenakker C W J 1991 *Phys. Rev.* **44** 1646
- [6] Kouwenhoven L P, van der Vaart N C, Johnson A T, Kool W, Harmans C J P M, Williamson J G, Staring A M and Foxon C T 1991 *Z. Phys. B* **85** 367
- [7] Hallam L D, Weis J and Maksym P A 1994 in preparation
- [8] Averin D V and Nazarov Yu V 1990 *Phys. Rev. Lett.* **65** 2446
- [9] McEuen P L, Foxman E B, Meirav U, Kastner M A, Meir Y, Wingreen N S and Wind S J 1991 *Phys. Rev. Lett.* **66** 1926
- [10] Ashoori R C, Störmer H L, Weiner J S, Pfeiffer L N, Pearton S J, Baldwin K W and West K W 1993 *Phys. Rev. Lett.* **71** 613
- [11] Weis J, Haug R J, von Klitzing K and Ploog K 1993 *Phys. Rev. Lett.* **71** 4019
- [12] Weinmann D, Häusler W, Pfaff W, Kramer B and Weiss U 1994 *Europhys. Lett.* **26** 467
Kramer B, Brandes T, Häusler W, Jauregui K, Pfaff W and Weinmann D 1994 *Semicond. Sci. Technol.* **9** 1871

Phase-breaking rates from conductance fluctuations in a quantum dot

C M Marcus[†], R M Clarke[†], I H Chan[†], C I Duruöz[‡] and J S Harris[‡]

[†] Department of Physics, Stanford University, Stanford, CA 94305-4060, USA

[‡] Solid State Laboratories, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

Abstract. We report measurements of universal-like conductance fluctuations in a quasiballistic $\sim 2 \mu\text{m}^2$ GaAs/AlGaAs quantum dot with adjustable point contact leads. Measurements cover a range of conductance G from nearly isolated, $\langle G \rangle \sim 0.1e^2/h$, to several modes in each lead, $\langle G \rangle \sim 5e^2/h$. The characteristic magnetic field scale of the fluctuations is found to increase with increasing mean conductance through the dot, consistent with a semiclassical picture of escape through the leads, and provides a means of estimating the phase breaking time τ_ϕ for electrons inside the dot. The fluctuation amplitude is found to increase with increasing dot conductance, then begins to saturate once a few conducting channels are open.

1. Introduction

Mesoscopic effects such as universal conductance fluctuations (UCF) and weak localization appear when electron phase is preserved during transport, giving rise to quantum interference from multiple conduction paths through the sample [1]. The characteristic time over which phase coherence is lost, for instance due to electron–phonon or electron–electron scattering, defines a phase breaking time τ_ϕ (and a rate $\gamma_\phi = 1/\tau_\phi$). The physical mechanism of phase breaking and its dependence on temperature and sample dimensionality have been widely investigated theoretically [2, 3] and experimentally [4–7] in pure and disordered conductors. In disordered metals and semiconductors, the elastic scattering time τ is much shorter than τ_ϕ , so that phase-coherent electrons move diffusively in at least one direction through the sample. This is the situation usually considered in mesoscopic systems. By contrast, in high-mobility two-dimensional electron gases (2DEGS) inelastic scattering may be the dominant scattering rate, so that electrons move ballistically before suffering phase breaking scattering. Diffusive versus ballistic motion distinguishes two regimes of coherent transport which have distinct properties, including phase breaking rates [7].

The confinement of electrons to reduced dimension also affects phase breaking rates. Dimensional crossover occurs when the lateral dimension L of the sample becomes smaller than the phase breaking length, $L_\phi = \sqrt{D\tau_\phi}$ (D is the diffusion constant) for diffusive transport, or $L_\phi = v_F\tau_\phi$ for ballistic transport. In this paper, we address low-temperature phase breaking in a micro-metre-size GaAs/AlGaAs quantum dot with two point contact leads (figure 1). By the definition above, the dot is zero dimensional (0D) since $L_\phi > L$ at low temper-

atures. Transport through the dot is also quasiballistic, by which we mean that all lateral dimensions are larger than or comparable to the elastic mean free path, $L \gtrsim l = v_F\tau$. To our knowledge, no theory of phase breaking in 0D structures has appeared in either the diffusive ($l \ll L$) or ballistic ($l > L$) regimes. (However, new results for diffusive 0D structures were presented by Imry [8] at the conference associated with these proceedings.) Experimentally, several groups have investigated phase breaking in quasiballistic GaAs quantum wires [9–11]. In each case, values for τ_ϕ were extracted from transport data using various weak localization or UCF formulae, all of which assume diffusive transport. Phase breaking rates found in [9–11] appear to saturate for $T \lesssim 0.5$ K, which may be due to electron heating but may also be associated with ballistic transport.

In order to develop a means of extracting τ_ϕ from transport measurements in 0D and nondiffusive structures, Marcus *et al* [12] proposed a simple extension of the semiclassical analysis of ballistic conductance fluctuations by Jalabert *et al* [13] by including a phase breaking term. In the present paper, we confirm experimentally a prediction of that analysis which relates the characteristic magnetic field of the conductance fluctuations to the average conductance through the dot. We will then apply these results and extract a value for τ_ϕ at $T = 20$ mK. The value we find is consistent with [9–11] and gives an order of magnitude estimate $\tau_\phi \sim \hbar/kT$ at that temperature.

2. Measurements

A micrograph of the quantum dot we measured is shown in figure 1. The device was formed using electrostatic gates (150 Å Ti, 400 Å Au) on the surface of a GaAs/

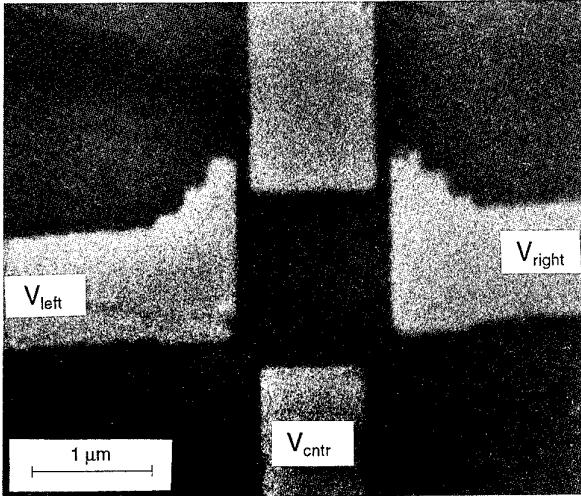


Figure 1. Electron micrograph of a quantum dot. Light regions are confining gates, dark regions are the GaAs surface. The small gaps separating the upper gate from the side gates are depleted during operation and do not provide exit paths; the larger gaps at the lower left and lower right corners form the two leads of the device.

$\text{Al}_x\text{Ga}_{1-x}\text{As}$ heterostructure, 420 Å above the 2DEG layer, patterned using electron-beam lithography. The lithographic dimensions define a square device with side length 1.5 μm and 260 nm point contacts on adjacent corners. Assuming a depletion width of ~50 nm between the edge of the gate and the 2DEG gives an approximate effective dot area $A_{\text{dot}} \sim 2 \mu\text{m}^2$. Mobility and sheet density were measured to be $\mu = 176\,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $n_s = 3.6 \times 10^{11} \text{ cm}^{-2}$ at 20 mK in an ungated Hall bar made from the same wafer. Gate voltages V_{left} and V_{right} were set so that the resistance across the dot was twice the value obtained when the other was set to zero. This provided roughly equal conductances in the two leads. Once V_{left} and V_{right} were set in this way, the overall conductance of the dot was changed using the centre gate voltage V_{cntr} . Sample resistance ($= V/I_{\text{bias}}$) was measured as a function of applied perpendicular magnetic field using standard four-wire AC lock-in techniques with 1 nA current bias at 11 Hz.

Two magnetic field scales can be identified for this device: a classical characteristic field $B_{\text{cl}} \sim 71 \text{ mT}$ defined by the condition that the cyclotron diameter equal the side length, $2r_{\text{cyc}}(B) = L$, and a quantum characteristic field $B_{\text{q}} \sim 2 \text{ mT}$ defined by the condition $A_{\text{dot}} B_{\text{q}} = \phi_0$, i.e. one flux quantum ($\phi_0 = h/e$) through the area of the dot. Typical low-field conductance fluctuations for three values of V_{cntr} are shown in figures 2(a), (c), (e).

3. Analysis

The semiclassical theory of conductance fluctuations in ballistic and quasiballistic microstructures assumes that the electron trajectories within the device are chaotic, leading to an exponential escape probability $P(t) \propto$

$\exp(-\gamma_{\text{esc}} t)$ for an electron to remain inside the dot for a time t before exiting via one of the leads [13, 14]. The escape rate γ_{esc} can be related to dynamical properties of the chaotic repeller inside the dot [15] and also to the total transmission of all exit channels by the expression

$$\Sigma_c T_c = 2\pi \frac{\hbar \gamma_{\text{esc}}}{\Delta} \quad (1)$$

where T_c is the total transmission of the c th lead (T_c can be greater than 1 for a multichannel lead) and Δ is the mean spacing of spin degenerate energy levels, $\Delta \simeq 2\pi\hbar^2/m^* A_{\text{dot}}$, ignoring perimeter corrections [16]. In [12], phase breaking was accounted for by introducing an effective number of escape channels with total transmission T_ϕ into which coherent electrons could escape. These phase breaking channels do not source or sink any current, but will affect fluctuation statistics. We emphasize that the phase breaking channels are fictitious; there are no extra physical voltage probes. This description merely represents the assumption that escape and phase breaking rates may be added. Within this picture, the effective total transmission out of the dot for coherent electrons is proportional to the sum of rates $\gamma_{\text{esc}} + \gamma_\phi$ in analogy with equation (1)

$$\Sigma_c T_c + T_\phi = 2\pi \frac{\hbar(\gamma_{\text{esc}} + \gamma_\phi)}{\Delta}. \quad (2)$$

Chaotic scattering also gives rise to an exponential distribution for the signed area, A , that is swept out by trajectories before escape, $P(A) \propto e^{-2\pi\alpha|A|}$. It is the distribution $P(A)$, characterized by the parameter α (an inverse area) which determines the spectrum of magnetoconductance fluctuations. This is because phase shifts induced by an applied magnetic field B are proportional to the Aharonov–Bohm flux AB (neglecting classical curvature). From this distribution, Jalabert *et al* [13] calculated a Lorentzian-squared form for the semiclassical autocorrelation of conductance fluctuations $C(\Delta B) = \langle \delta g(B + \Delta B) \delta g(B) \rangle_B$ based on a Landauer formula relating conductance to the transmission matrix. The Fourier transform of $C(\Delta B)$ gives the power spectrum $S(f)$ of conductance fluctuations in terms of the characteristic inverse area α

$$S(f) = \text{FT}(C(\Delta B)) = S(0)(1 + 2\pi\alpha\phi_0 f) \exp(-2\pi\alpha\phi_0 f), \quad (3)$$

where f is the magnetic frequency in units of cycles/tesla, and $\phi_0 = h/e$ (= 4.14 mT μm²) is the quantum of magnetic flux. A semiclassical treatment is expected to be valid in the region of overlapping resonances, $\hbar\gamma \gg \Delta$, which implies that the characteristic dwell time $1/\gamma$ is much less than the Heisenberg time $\tau_H = \hbar/\Delta$ beyond which individual quantum levels become resolvable. From equation (2), this requirement is equivalent to requiring many conducting channels, $\Sigma_c T_c + T_\phi \gg 1$.

The accumulation of swept area by a chaotic trajectory occurs as a random walk so that escape area is proportional to the square root of escape time, hence $\alpha \propto \sqrt{\gamma}$ with a constant of proportionality that depends

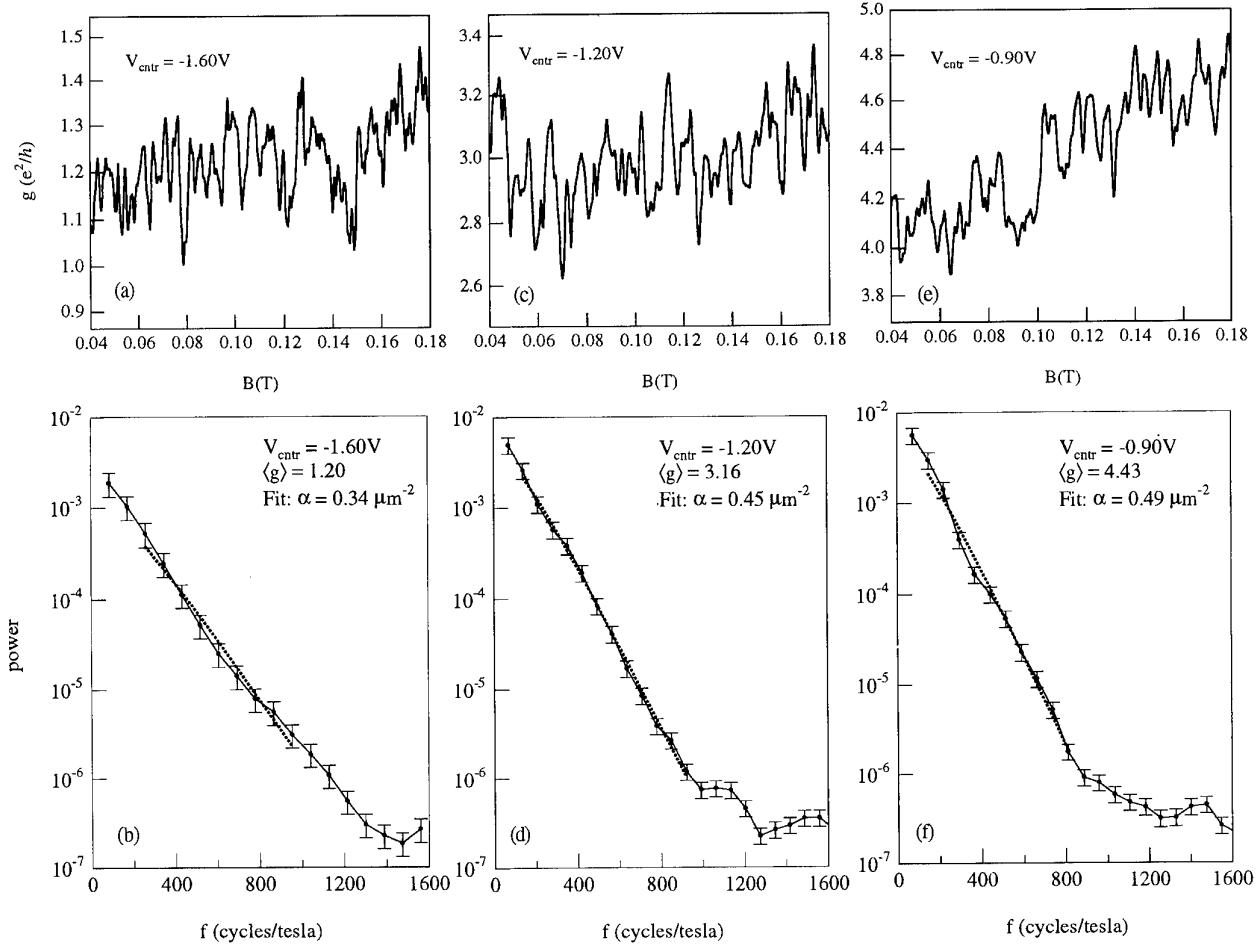


Figure 2. Magnetoconductance fluctuations with corresponding average power spectra for three values of mean conductance: (a, b) $\langle g \rangle \sim 1.2$; (c, d) $\langle g \rangle \sim 3.2$; (e, f) $\langle g \rangle \sim 4.4$, at a mixing chamber temperature of $T \sim 20$ mK. Decreasing characteristic magnetic field of the fluctuations with decreasing $\langle g \rangle$ is visible in the conductance data. Power spectra below each trace include fits to equation (3) (dotted curves) from which characteristic inverse areas α are extracted.

on the geometry of the dot [17]. Because only coherent electrons contribute to conductance fluctuations, the relevant γ in this expression should include phase breaking. From equation (2), one therefore expects that $\alpha^2 \propto (\Sigma_c T_c + T_\phi)$. Intuitively, this proportionality reflects the fact that the distribution of areas detected in an interference measurement should only include electrons that have not ‘escaped’ either via the leads (the $\Sigma_c T_c$ part) or via phase breaking (the T_ϕ part). For a two-lead geometry ($c = 1, 2$), the sum over leads can be further related to the average conductance through the dot. If we assume that electrons become ‘well mixed’ within the structure, then the average conductance (averaged over fluctuations, i.e. an appropriate range of magnetic field) is given by the resistors-in-series form $\langle g \rangle = 2T_1 T_2 / (T_1 + T_2)$, where g is in units of e^2/h and the factor of 2 accounts for spin. Notice that T_ϕ does not appear in this expression. For simplicity, we will assume $T_1 = T_2$ (the approximate experimental situation), in which case $\Sigma_c T_c = 2\langle g \rangle$, but all results can be easily generalized to $T_1 \neq T_2$. One then expects a scaling between the mean conductance $\langle g \rangle$ and the characteristic inverse area α of

the form

$$\alpha^2 = k \left[\langle g \rangle + \left(\frac{\pi \hbar}{\Delta} \right) \gamma_\phi \right] \quad (4)$$

with some geometry-dependent constant k . This expression allows values for γ_ϕ to be extracted from fits to the power spectrum of ballistic conductance fluctuations.

4. Results

Figure 2 shows three representative sets of conductance fluctuation data for different mean conductances, $\langle g \rangle \sim 1.2, 3.2$ and 4.4 . Below each trace is an averaged power spectrum $S_g(f)$ of the conductance fluctuations $\delta g(B)$, computed over the field range shown. The fluctuations $\delta g(B)$ were extracted from the full $g(B)$ by subtracting a cubic polynomial that was fitted over the region of interest. The spectra are roughly exponential down to the noise floor, and a logarithmic least-squares fit to the form in equation (3) over the range $200 < f < 800$ cycles/tesla can be used to extract values for α . We note that

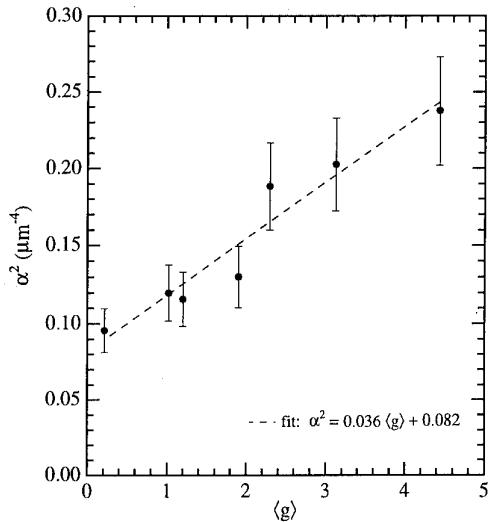


Figure 3. Plot of α^2 versus mean conductance $\langle g \rangle$ in the field range $0.04 \text{ T} < |B| < 0.2 \text{ T}$. Conductance was controlled by changing the gate voltage V_{ctr} . The broken line is a fit to the form $\alpha^2 = A\langle g \rangle + B$, suggested by equation (4).

at low $\langle g \rangle$ the semiclassical analysis leading to equation (3) is not strictly valid, and in fact it appears that the lowest conductance spectrum (figure 2(b)) agrees less well with the form of equation (3) and has a more concave-up shape than either the theoretical curve or the higher-conductance spectra. This characteristic deviation towards greater high-frequency content is consistent with theoretical [18] and numerical [19] studies of chaotic scattering beyond the semiclassical regime.

Figure 3 shows α^2 versus $\langle g \rangle$ for several centre gate voltages. The data (full circles) are in reasonable agreement with the prediction of our model, equation (4). A least-squares fit gives $\alpha^2 = 0.036\langle g \rangle + 0.082$. Using these values and $\Delta = 3.6 \mu\text{eV}$, we find a phase breaking rate $\gamma_\phi = \Delta/\pi\hbar(0.082/0.036) = 3.9 \times 10^9 \text{ s}^{-1}$, or $\tau_\phi = 260 \text{ ps}$. The uncertainty is rather large, around 20%, due to uncertainties in the device area (which determines Δ), the value of α extracted from the power spectra, and the fit to α^2 versus $\langle g \rangle$. Expressing γ_ϕ in terms of the number of phase breaking channels gives $T_\phi = 2(0.082/0.036) = 4.6 \pm 1$, and in terms of a temperature gives $\hbar\gamma_\phi/k \sim 30 \text{ mK}$, comparable to the mixing chamber temperature $T \sim 20 \text{ mK}$. The temperature dependence of γ_ϕ will be reported in a subsequent paper; however, preliminary results reported in [12], and in the present experiment as well, indicate the expected trend, namely as the temperature is raised, α increases for fixed $\langle g \rangle$. The value of τ_ϕ also appears roughly consistent with a number of recent measurements of phase breaking in 1D diffusive samples [9–11]. Consistent with these experiments, our value for τ_ϕ is smaller than the predicted Nyquist time [2] associated with electron-electron scattering, the expected dominant phase-breaking mechanism at this temperature in diffusive samples. We emphasize, however, that existing theory explicitly does not apply to quantum dots, where $L \sim l \ll L_\phi$.

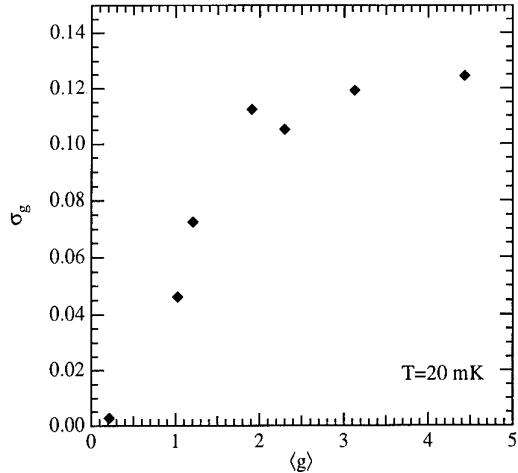


Figure 4. Standard deviation of conductance fluctuations, $\sigma_g = \sqrt{\text{var}(\delta g(B))}$, versus mean conductance $\langle g \rangle$ in the field range $0.04 \text{ T} < |B| < 0.2 \text{ T}$. Conductance was controlled by changing the gate voltage V_{ctr} .

We have also investigated the dependence of fluctuation amplitude, $\sigma_g = \sqrt{\text{var}(\delta g(B))}$, on mean conductance at low fields, $0.04 \text{ T} > B > 0.2 \text{ T}$. Figure 4 shows that at low conductance, $\langle g \rangle < 1$, the fluctuation amplitude increases with $\langle g \rangle$, but appears to saturate near a value $\sigma_g \sim 0.12$ once several conducting channels are open. Berry *et al* [20] recently reported comparable fluctuation amplitudes, but within a strong saturation of σ_g in a stadium-shaped billiard with a factor of two higher mobility. The numerical results of Baranger *et al* [19] show a saturation of fluctuation amplitude with the number of conducting channels at a value $\sigma_g \sim 0.25$ for asymmetric chaotic-shaped structures, but the integrable shapes studied did not show this saturation. In either case, the numerical data in [19] did not reveal the sharp decrease in the fluctuation amplitude for small $\langle g \rangle$ which we observe experimentally. The difference may be attributable to phase breaking or thermal smearing of fluctuations in the experiment. Iida *et al* [21] studied a 1D diffusive model that included a third lead (voltage probe) to provide phase breaking, and found that strong coupling to the third lead decreases the fluctuation amplitude. A simple theory of σ_g versus $\langle g \rangle$ based on a Hauser–Feshbach-type approach [12] predicts a fluctuation amplitude that depends on phase breaking rate according to $\sigma_g = \langle g \rangle^2 / (\langle g \rangle + T_\phi/2)$ for the case of equivalent, single-channel leads (this implies $\langle g \rangle < 1$). The data in figure 4 are in good qualitative agreement with this expression for small $\langle g \rangle$, but do not agree quantitatively if we insert the value $T_\phi = 4.6$. This discrepancy remains a puzzle, and will be investigated more fully in the future.

After completing this paper, we received a preprint from Baranger and Mello [22] showing analytically and numerically that conductance fluctuations in a quantum dot at $T = 0$, i.e. without thermal smearing or phase breaking, have a universal amplitude (in our notation, $\sigma_g = 0.5$ in a magnetic field) as the number of channels N

in each lead becomes large. Furthermore the amplitude is rather insensitive to N , and slightly increases as N decreases to 1. The disagreement of these results with present and previous experiments is probably due to finite temperature effects in the experiments.

Acknowledgments

We thank Chris Knorr at Stanford's Center for Integrated Systems for assistance in sample fabrication, and H Baranger, Y Imry, M Keller and C Lewenkopf for valuable discussions. Research was supported in part by the NSF-MRL through the Stanford Centre for Materials Research, and by JSEP under contract DAAL03-91-C-0010.

References

- [1] Altshuler B L, Lee P A and Webb R (eds) 1991 *Mesoscopic Phenomena in Solids* (Amsterdam: North-Holland)
- [2] Altshuler B L, Aronov A G and Khmel'nitsky D E 1982 *J. Phys. C: Solid State Phys.* **15** 7367
- [3] Stern A, Aharonov Y and Imry Y 1990 *Phys. Rev. A* **41** 3436
- [4] Echternach P M, Gershenson M E and Bozler H M 1993 *Phys. Rev. B* **48** 11516
- [5] Choi K K, Tsui D C and Alavi K 1987 *Phys. Rev. B* **36** 7551
- [6] Washburn S and Webb R A 1993 *Rep. Prog. Phys.* **55** 1311
- [7] Yacoby A, Sivan U, Umbach C P and Hong J M 1991 *Phys. Rev. Lett.* **66** 1938
- [8] Imry Y and Stern A 1994 *Semicond. Sci. Technol.* **9** 1879
- [9] Ikoma T and Hiramoto T 1991 *Granular Nanoelectronics* ed D K Ferry (New York: Plenum)
- [10] Kurdak Ç, Chang A M, Chin A and Chang T Y 1992 *Phys. Rev. B* **46** 6846
- [11] Bird J P, Grassie A D C, Lakrimi M, Hutchings K M, Harris J J and Foxon C T 1990 *J. Phys.: Condens. Matter* **2** 7847
- [12] Marcus C M, Westervelt R M, Hopkins P F and Gossard A C 1993 *Phys. Rev. B* **48** 2460
- [13] Jalabert R A, Baranger H U and Stone A D 1990 *Phys. Rev. Lett.* **65** 2442
- [14] Doron E, Smilansky U and Frenkel A 1991 *Physica D* **50** 367
- [15] Smilansky U 1990 *Chaos and Quantum Physics* ed M Giannoni *et al* (Amsterdam: Elsevier) p 371
- [16] Doron E and Smilansky U 1992 *Phys. Rev. Lett.* **68** 1255
- [17] Jensen R V 1991 *Chaos* **1** 101
- [18] Lewenkopf C H and Weidenmüller H A 1991 *Ann. Phys. NY* **212** 53
- [19] Baranger H U, Jalabert R A and Stone A D 1993 *Chaos* **3** 665
- [20] Berry M J, Katine J A, Marcus C M, Westervelt R M and Gossard A C 1994 *Surf. Sci.* **305** 495
- [21] Iida S 1991 *Phys. Rev. B* **43** 6459
- [22] Baranger H U and Mello P A 1994 *Phys. Rev. Lett.* **73** 142

Chaotic transport and fractal spectra in lateral superlattices

R Fleischmann, T Geisel, R Ketzmerick and G Petschelt†

Institut für Theoretische Physik und SFB Nichtlineare Dynamik, Universität Frankfurt, Postfach 11 19 32, D-60054 Frankfurt/Main, Germany

Abstract. We show that chaos and nonlinear resonances are clearly reflected in the magnetotransport of lateral surface superlattices and thereby explain a series of magnetoresistance peaks observed in antidot arrays on semiconductor heterojunctions. For small magnetic fields we find the counterintuitive result that electrons move in the *opposite* direction to the free-electron $E \times B$ drift when subject to a two-dimensional periodic potential. We show that this phenomenon arises from chaotic channelling trajectories, and by a subtle mechanism leads to a *negative* value of the Hall resistivity for small magnetic fields. For a quantum mechanical description of Bloch electrons in magnetic fields Harper's equation has been studied extensively; this is *integrable* in the classical limit and thus fails for lateral surface superlattices where *chaotic* trajectories prevail near the classical limit. We therefore derive a new model, which is exact under the most general conditions, study the influence of classical chaos on the fractal spectrum known as Hofstadter's butterfly, and make predictions on its observability in lateral surface superlattices.

1. Introduction

As the trend to smaller and faster microstructures progresses, understanding the nonlinear dynamics of electrons in anharmonic potentials becomes increasingly important. At the present stage, where the Fermi wavelength is still smaller than typical spatial scales, much of the dynamics can be described by chaotic classical trajectories, but in the future quantum mechanical treatments in the semiclassical regime will be required.

We will give an overview of recent work on applications of nonlinear dynamics to antidot lattices and will demonstrate how fundamental concepts like the KAM theorem are manifested in recent experimental results [1–3]. In particular it was shown that a series of peaks in the magnetoresistance can be explained entirely by chaotic trajectories, as regular trajectories remain pinned in an electric field due to the KAM theorem [4]. The negative Hall effect in antidot lattices was explained by chaotic channelling trajectories [5]. Similarly the occurrence of several non-quantized Hall plateaus in ballistic microjunctions was explained by nonlinear dynamics [6].

A quantum mechanical description has to tackle the problem of Bloch electrons in magnetic fields. In the framework of Harper's equation it was shown that the dynamics of wavepackets exhibits an unusual diffusive spread [7] and that correlation functions decay

extremely slowly in the form of a power law [8]. Both properties can be related to multifractal properties of the energy spectrum. Quantum magnetotransport properties have been calculated by diagonalizing the Hamiltonian of equation (1) using magnetic Bloch functions [9]. The influence of classical chaos on a fractal spectrum was studied in a newly developed vectorial tight-binding equation [10], which was derived without any approximations from the single-particle Hamiltonian of equation (1). It thus overcomes the serious failures of Harper's equation for lateral superlattices and lends itself to applications in experimental situations.

2. Chaotic dynamics and magnetotransport in antidot arrays

The classical approximation for the dynamics of an electron wavepacket in a two-dimensional potential $U(x, y)$ and a perpendicular magnetic field B is described by the Hamiltonian

$$H = (\mathbf{p} - e\mathbf{A})^2/2m + U(x, y) \quad (1)$$

where \mathbf{A} is the vector potential and m the effective mass of the electron [11]. To model antidot arrays we will use the potential (see figure 1(a))

$$U(x, y) = U_0 [\cos(\pi x) \cos(\pi y)]^\beta \quad (2)$$

where β controls the steepness of the antidots and the prefactor U_0 of the potential is chosen such that the ratio of the dot diameter at the Fermi energy to the

† Present address: Potsdam-Institut für Klimafolgenforschung, Telegrafenberg, D-14473 Potsdam, Germany.

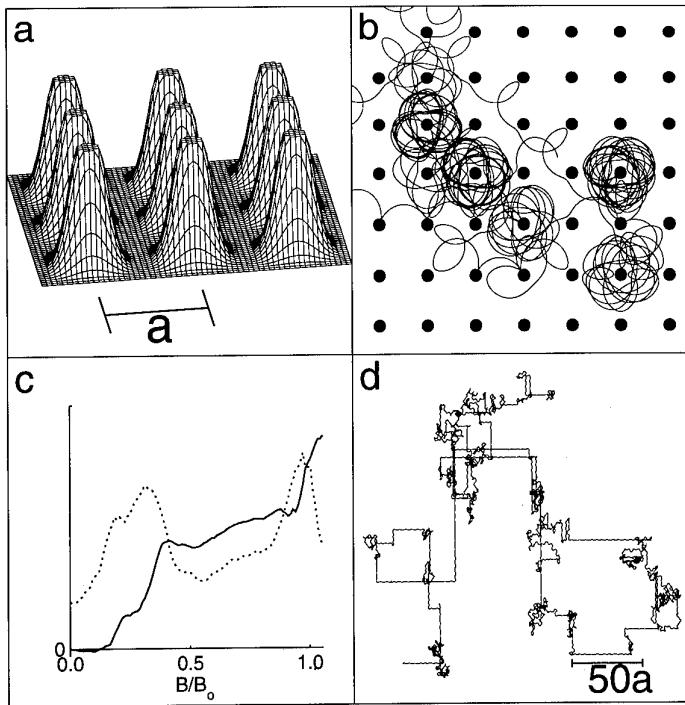


Figure 1. (a) The antidot potential (equation (2)) for $\beta = 4$.
(b) Chaotic electron trajectory that is trapped several times around an antidot. (c) Magnetoresistance ρ_{xy} (dotted curve) showing peaks at magnetic fields where chaotic electrons are often trapped around 1 and 4 antidots, and Hall resistivity ρ_{xy} (full curve) showing a negative value for small magnetic fields. (d) Chaotic electron trajectory for small magnetic fields ($B/B_0 = 0.1$) where the negative Hall effect occurs (note the much larger scale than in (b)).

separation of adjacent dots is one third, similar to that in the experiments [1].

For certain magnetic field regimes we find trajectories enclosing 1, 2, 4, 9, 21, and even more antidots for a steep potential ($\beta = 64$) and orbits enclosing 1 and 4 antidots for a smooth potential ($\beta = 4$) as well as chaotic orbits, as shown in figure 1(b). In order to calculate transport properties one would like to use linear response theory, e.g. Kubo's formula, but this turns out to be incorrect for those regular trajectories that surround some number of antidots. The reason lies in a general property of chaotic systems given by the KAM theorem [12]. These regular trajectories correspond to nonlinear resonances in phase space which remain stable under a small perturbation, e.g. an applied electric field. Therefore they are pinned around the antidots even when an electric field is applied, and they do not contribute to the magnetotransport at all. The transport properties of the remaining chaotic trajectories can be described using Kubo's formula, and we find for the total frequency-dependent conductivity including impurity scattering with an average time interval τ between scattering events [4]

$$\sigma_{ij}(\omega, \tau) = (1 - p_r) \frac{ne^2}{k_B T} \int_0^\infty dt e^{(-1/\tau + i\omega)t} \langle v_i(t)v_j(0) \rangle \quad (3)$$

where the correlation function includes only unperturbed chaotic orbits and p_r is the phase space fraction of the regular pinned orbits.

With this model we were able to reproduce the experimental results for the magnetoresistance of Weiss *et al* [1] and in addition to give a much more detailed explanation. It turns out that the magnetic field dependence of the phase space fraction p_r of regular orbits in the prefactor of equation (3) is far too weak to account for the pronounced magnetoresistance peaks shown in figure 1(c). They are caused mainly by chaotic orbits, like the one in figure 1(b), that are trapped around 1 or 4 antidots for long times. This trapping can be explained by so-called cantori in phase space close to the nonlinear resonance of regular motion [13]. These cantori occur in a hierarchy that can trap a trajectory for an algebraically long time and are a generic phenomenon of chaotic systems with mixed (i.e. regular and chaotic) phase space.

3. Negative Hall effect

For small magnetic fields (cyclotron radius larger than five unit cells) the electron dynamics in an antidot

array shows a surprising feature (figure 1(d)). Chaotic electron trajectories tend to stay within the channels between the antidots along the axes for 3 to 1000 or even more unit cells. This again is due to the cantori in phase space around a nonlinear resonance [5]. The Hall conductivity is determined, as in equation (3), by the correlation function $\langle v_x(t)v_y(0) \rangle$, which we found to be dominated by the long channelling episodes of chaotic trajectories and the way they eventually leave the channel. Because of the magnetic field the channelling trajectories tend to move on one side of the channel and thus they have an increased probability of hitting one of the antidots on this side such that they leave the channel to the opposite side and thereby turn against the Lorentz force. The combination of the nonlinear dynamics that causes chaotic orbits to stay on one side of the channel and the above-mentioned geometrical effect when leaving the channel cause the correlation function and the Hall resistivity (figure 1(c)) to be negative [5]. This explains the quenched and negative Hall effect in recent experiments on antidot superlattices [1, 3].

4. Bloch electrons in a magnetic field

The description of electrons in a periodic potential and a competing magnetic field has posed theoretical problems for a long time. As a hypothesis Peierls suggested solving the Schrödinger equation for $B = 0$, choosing one of the bands produced by the periodic potential and substituting the Bloch index k by the operator $(p - eA)/\hbar$ in the dispersion $\varepsilon(k)$ [14]. Later authors have worked on justifications of this hypothesis [15]. Hence, assuming a tight-binding dispersion $\varepsilon(k)$, one obtains Harper's equation [16], which was shown by Azbel and Hofstadter to exhibit a self-similar spectrum for extremely large magnetic fields [17–19]. Considering the approximations involved (single tight-binding band, Peierls hypothesis for $B \rightarrow 0$), one may wonder how much of this spectrum, known as Hofstadter's butterfly, survives in a physically realistic model. Also, in a dual Harper's equation derived by Rauh [20] and Langbein [21] in the opposite limit $B \rightarrow \infty$ the quality of predictions for realistic systems remains unclear.

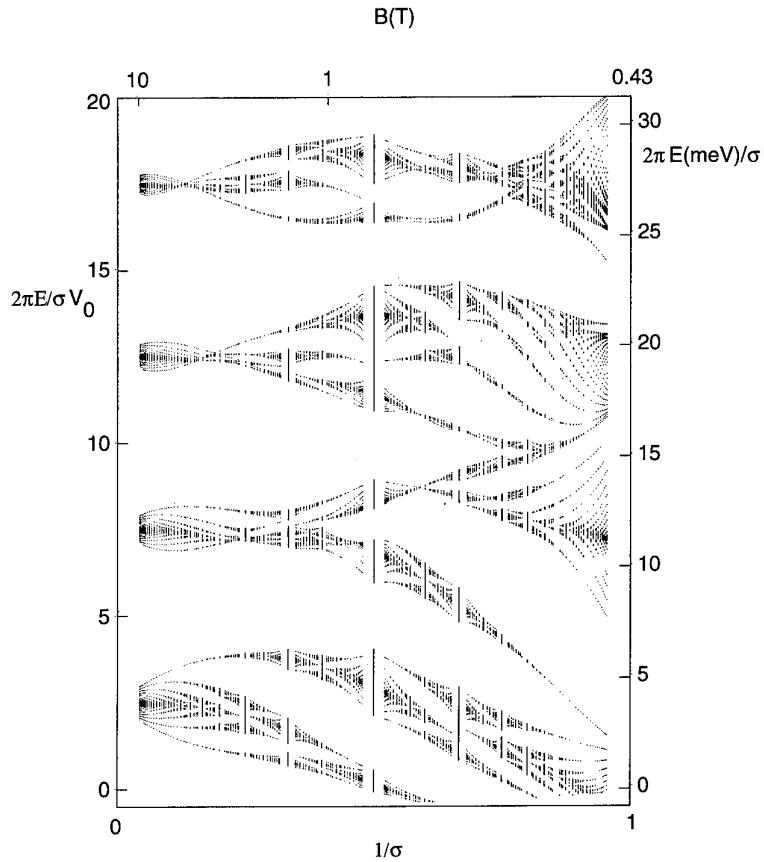


Figure 2. Band structure of the four lowest Landau bands as a function of $1/\sigma$ ($\sigma = eBa^2/h$ is the number of flux quanta per unit cell a^2) at $\Omega_0 = h^2/mV_0a^2 = 5.0$. In order to remove the B -dependence of the Landau levels the energy is multiplied by $2\pi/\sigma$. The bottom and left axes are given in dimensionless units $1/\sigma$ and $2\pi E/\sigma V_0$, while the top and right axes refer to a lattice constant of $a = 100$ nm and are labelled in absolute magnetic field and energy units. For large values of the magnetic field the bands are well separated and still exhibit structures similar to the original Hofstadter butterfly. At smaller fields, however, this is no longer the case as the bands overlap strongly and lose their self-similarity.

As the detection of Hofstadter's butterfly in natural crystals would require magnetic fields of about 10^5 T, one now tries to achieve it in lateral surface superlattices (LSSLs) on semiconductor heterojunctions where, due to a larger unit cell a^2 , the interesting regime of about one flux quantum per unit cell can be obtained by smaller values of B . We pointed out previously [7, 22] that one thereby approaches the classical limit ($a \gg \lambda_F$) where chaotic trajectories prevail, as manifested e.g. in magnetoresistance experiments [1–5]. Harper's equation, however, belongs to an integrable Hamiltonian and thus misses chaotic behaviour in its classical limit. Strictly speaking, Hofstadter's butterfly no longer applies to LSSLs (i.e. in near classical systems), while on the other hand experimentalists try to detect it in LSSLs for reasons of accessible magnetic fields. A new realistic model is required.

In reference [10] we derived a vectorial generalization of Harper's equation from the Hamiltonian of equation (1) without using approximations, and showed that it holds under most general conditions. It is exact and brings forth chaotic behaviour in its classical limit as it should. It is thus applicable to quantum systems in the semiclassical regime like LSSLs. It allows us to investigate the influence of classical chaos on Cantor spectra [7, 22] (i.e. uncountable fractal spectra) in a physically realistic model, a problem that previously was studied intensely for discrete spectra.

In particular, using a matrix continued-fraction expansion for the Green function, we determined Hofstadter's butterfly under realistic conditions. Figure 2 shows that for strong magnetic fields the bands are well separated and butterfly look-alikes emerge. For decreasing magnetic fields the Landau bands overlap, as is seen most typically in the mixing between the second and third Landau bands. The fourth band is well isolated again, but it rather resembles a fish than a butterfly. The upper and right axes give the spectrum in absolute units for the particular choice of $a = 100$ nm.

For the observation of butterfly structures in experiments on LSSLs it is a serious problem to avoid strongly overlapping bands, which destroy the self-similarity. Furthermore one wants to resolve energy gaps at least in the second hierarchical step of the fractal spectrum. These gaps must have a width of at least 0.1 meV to override the broadening of levels due to impurity scattering. As the energy scale in figure 2 is given in units of the potential V_0 an increase of the modulation strength would result in an increase of the gap widths. On the other hand this would also lead to stronger band mixing and the regime of non-overlapping bands would be shifted to even larger magnetic fields. The latter can be avoided by appropriately decreasing the lattice period a . Assuming the scaling properties of the original Hofstadter butterfly, the above gap width of 0.1 meV would require a modulation strength of about 5–10 meV. Thus for $\Omega_0 = h^2/mV_0a^2 \approx 4.0$ the observability of butterfly structures in a sufficiently

wide magnetic field regime down to 1.5 T would require lattice constants $a \leq 70$ nm. These absolute values can easily be scaled for other lattice parameters and magnetic fields. For the first time this is a quantitative prediction for butterfly look-alikes in experiments on lateral superlattices.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft. RK thanks Walter Kohn for the hospitality in Santa Barbara and for partial support by the NSF under DMR90-01502.

References

- [1] Weiss D, Roukes M L, Menschig A, Grambow P, von Klitzing K and Weimann G 1991 *Phys. Rev. Lett.* **66** 2790
- [2] Lorce A, Kotthaus J P and Ploog K 1991 *Phys. Rev. B* **44** 3447
- [3] Schuster R, Ensslin K, Kotthaus J P, Holland M and Beaumont S P 1992 *Superlatt. Microstruct.* **12** 93
- [4] Fleischmann R, Geisel T and Ketzmerick R 1992 *Phys. Rev. Lett.* **68** 1367
- [5] Fleischmann R, Geisel T and Ketzmerick R 1994 *Europhys. Lett.* **25** 219
- [6] Geisel T, Ketzmerick R and Schedletzky O 1992 *Phys. Rev. Lett.* **69** 1680
- [7] Geisel T, Ketzmerick R and Petschel G 1991 *Phys. Rev. Lett.* **66** 1651
- [8] Ketzmerick R, Petschel G and Geisel T 1992 *Phys. Rev. Lett.* **69** 695
- [9] Silberbauer H 1992 *J. Phys.: Condens. Matter* **4** 7355
Silberbauer H and Rössler U 1994 to be published
- [10] Petschel G and Geisel T 1993 *Phys. Rev. Lett.* **71** 239
- [11] Geisel T, Wagenhuber J, Niebauer P and Obermair G 1990 *Phys. Rev. Lett.* **64** 1581
Wagenhuber J, Geisel T, Niebauer P and Obermair G 1992 *Phys. Rev. B* **45** 4372
- [12] See e.g. Arnold V I 1989 *Mathematical Methods of Classical Mechanics* (New York: Springer)
- [13] Geisel T, Zacherl A and Radons G 1987 *Phys. Rev. Lett.* **59** 2503; 1988 *Z. Phys. B* **71** 117
Zacherl A, Geisel T, Nierwetberg J and Radons G 1986 *Phys. Lett.* **114** 317
- [14] Peierls R 1933 *Z. Phys.* **80** 763
- [15] Kohn W 1959 *Phys. Rev.* **115** 1460
Blount E I 1962 *Phys. Rev.* **126** 1636
Roth L 1962 *J. Phys. Chem. Solids* **23** 433
Brown E 1968 *Phys. Rev.* **166** 626
Zak J 1968 *Phys. Rev.* **168** 686
- [16] Harper P G 1955 *Proc. R. Soc. A* **68** 874
- [17] Azbel' M Ya 1964 *Sov. Phys.-JETP* **19** 634
- [18] Hofstadter D R 1976 *Phys. Rev. B* **14** 2239
- [19] For reviews see e.g. Sokolov J B 1985 *Phys. Rep.* **126** 189
Simon B 1982 *Adv. Appl. Math.* **3** 463
Hiramoto H and Kohmoto M 1992 *Int. J. Mod. Phys. B* **6** 281
- [20] Rauh A 1975 *Phys. Status Solidi b* **69** K9
- [21] Langbein D 1969 *Phys. Rev.* **180** 633
- [22] Geisel T, Ketzmerick R and Petschel G 1991 *Phys. Rev. Lett.* **67** 3635

Quantum transport and quantum chaos in antidot superlattices in a magnetic field

H Silberbauer, P Rotter, M Sührke and U Rössler

Institut für Theoretische Physik, Universität Regensburg, D-93040 Regensburg,
Germany

Abstract. The classical dynamics of electrons in antidot superlattices subject to a perpendicular magnetic field is known to be chaotic. The quantum aspects of this system are investigated by calculating the magnetotransport properties and by studying the level statistics. We obtain quantitative agreement with experimental data for the longitudinal and Hall resistances and find distributions of level separations which, depending on the magnetic wavevector, show level repulsion as a signature of quantum chaos.

1. Introduction

Antidot superlattices have become a popular system for investigating the complex dynamics of electrons under the influence of a periodic potential and a perpendicular magnetic field. In typical samples the superlattice periods are much larger than the Fermi wavelength of the electrons, so this dynamics is essentially in the classical regime. Magnetotransport measurements in clean samples with ballistic motion of the electrons between the antidots exhibit characteristic peaks in the longitudinal resistance at low magnetic fields. These peaks have been intuitively interpreted as being due to pinned classical orbits around one or more antidots if the cyclotron radius is commensurate with the lattice period [1]. Non-quantized plateaus in the Hall resistance are connected with these peaks and a dramatic quenching of the Hall effect has been observed for vanishing magnetic field. Similar observations have been made for rectangular and triangular geometries [2, 3].

A more detailed analysis of the classical dynamics reveals that the chaotic parts of the classical phase space are crucial for a more quantitative interpretation of the experimental findings. The commensurability peaks are now ascribed to chaotic trajectories near KAM tori on which electrons are trapped for a while around a group of antidots [4].

As in antidot superlattices the chaotic motion is nicely reflected in experiments, they seem to be an ideal system for the study of quantum manifestations of classical chaos. With a further reduction of the lattice constant from currently realized values (of about 150 nm) down to the order of the Fermi wavelength (of about 50 nm) one should be able to see the transition of a classical chaotic system to the quantum regime. This requires the application of a quantum transport theory [5] instead of the above-mentioned classical concepts.

One of the most frequently investigated quantum manifestations of classical chaos is the statistics of the

energy levels [6]. Quantum spectra typically show level repulsion if their classical analogue is fully chaotic. The observation of level repulsion in a given spectrum requires the avoidance of superposition of statistically independent subsets of the spectrum. Otherwise, the result would be level clustering as is also the case for classically regular systems [7]. In the case of level repulsion the distribution of level spacings vanishes for small spacings with a characteristic power. This power depends on the symmetry properties of the underlying Hamiltonian which can be divided into three universality classes [6]. A transition between different universality classes can be achieved, for example, by switching on a magnetic field, as it breaks time reversal symmetry [8]. There is continuous interest in the analysis of such transitions and in that from fully chaotic to regular systems which is characterized by a mixed classical phase space [9]. Different proposals have been made to describe the behaviour of the level distribution in this case [10–13] which have to be tested for their validity.

A system very close to an antidot superlattice is the Sinai billiard. This has been studied in [14] for periodic boundary conditions, i.e. on a closed domain which results in a discrete spectrum. In our case, however, one has to deal with an infinite lattice giving a continuous spectrum with energy bands. An example of such an unbounded system has been investigated in [15]. This model system, described by Harper's equation, is equivalent to a lateral superlattice in the limit of vanishing modulation potential. The authors investigated the distribution of energy gaps as an analogue to the distribution of spacings of discrete levels and found it to diverge for vanishing gap width due to the fractal properties of the spectrum for irrational flux through a unit cell.

In this paper it will be shown that quantum transport calculations of the longitudinal and Hall resistances for antidot superlattices yield results which are in quantitative agreement (even on absolute scales) with the

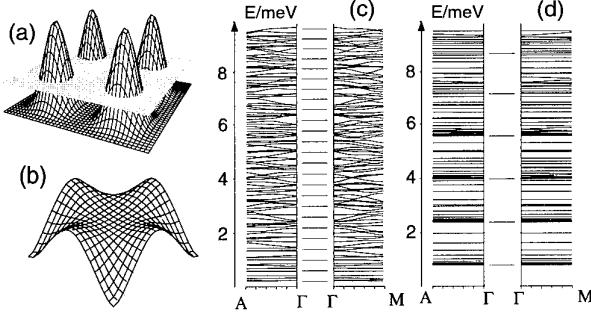


Figure 1. Shape of the antidot potential (a), a typical miniband dispersion in the MBZ (b), and dispersions for two different directions in the MBZ for magnetic fields $B = 0.23$ T (c) and $B = 0.92$ T (d) respectively. The potential chosen for the calculations in this paper has a height V_0 of 25 meV, a steepness parameter $\alpha = 4$ and a lattice constant a of 300 nm.

experimental data. We also perform level statistics of the energy spectra in our unbounded system characterized by a magnetic wavevector. Considering the energy spectra belonging to different wavevectors to be statistically independent we find level repulsion with different power laws as a signature of quantum chaos. With increasing magnetic field the system shows a transition to level clustering as the classical carrier dynamics becomes dominated by motion on undisturbed cyclotron orbits.

2. Quantum magnetotransport

In a first step we solve the eigenvalue problem of the single-particle effective-mass Hamiltonian for 2D electrons in a periodic potential $V(x, y)$ and a perpendicular magnetic field

$$H = \frac{1}{2m^*} \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 + V(x, y). \quad (1)$$

We choose the symmetric gauge for the vector potential $\mathbf{A} = B(-y, x, 0)/2$ and restrict ourselves to square lattices with lattice constant a . The shape of the phenomenological potential

$$V(x, y) = V_0 \left(\cos \frac{\pi x}{a} \cos \frac{\pi y}{a} \right)^{2\alpha} \quad (2)$$

is controlled by the parameters V_0 (for the height) and α (for the steepness). Thus, our model is just the quantum case of the classical model used in [4]. The parameters are chosen to model antidot potentials with potential maxima high above the Fermi energy E_F as shown in figure 1(a).

The eigenvalues of equation (1) are obtained by expanding the solutions in a series of magnetic Bloch functions as described in [16]. This concept is applicable only for so-called rational magnetic fields $B = r\Phi_0/a^2$ with $r \in \mathbb{P}$ and the magnetic flux quantum $\Phi_0 = hc/e$. For these magnetic fields the eigenstates $|n, \theta\rangle$ of H can

be classified by a magnetic wavevector θ and a miniband index n . As an example, we show for a potential with $V_0 = 25$ meV, $\alpha = 4$ and $a = 300$ nm (see figure 1(a)), a typical 2D dispersion of a single miniband in the magnetic Brillouin zone (MBZ) (figure 1(b)), and the dispersion of the 100 lowest minibands along two different directions in the MBZ for $B = 0.23$ T (figure 1(c)) and $B = 0.92$ T (figure 1(d)). The corresponding Landau ladders of the homogeneous 2D system are indicated between the miniband structures for comparison. In contrast to the weak modulation case [5], for which the miniband formation in general preserves the original Landau-level structure, the strong antidot potential, at least for lower magnetic fields, leads to a mixing of the Landau levels. Therefore, in this field region we find strong dispersion and overlapping minibands which are not reminiscent of the Landau ladder. For high magnetic fields the dispersion diminishes and the minibands cluster towards the Landau levels. There always remain, however, some edge states in the gaps between the Landau levels.

Transport properties at low temperatures depend not only on the properties of the pure quantum system, but are also strongly influenced by the scattering of the carriers by impurities. For samples with highly mobile carriers, as used in the underlying experiments, one may assume short-range impurity scattering which is here included within the well-established self-consistent Born approximation (SCBA). Thus we must solve the self-consistency equation [5]

$$\Sigma(z) = \gamma^2 \operatorname{tr}(G(z)) \quad (3)$$

for the self-energy, where $G(z)$ denotes the impurity-averaged Green operator. Here the self-energy is assumed to be independent of the quantum numbers n and θ and leads to a level broadening, described by γ , which can be derived in a standard way from the zero-field mobility of the homogeneous system which is known from experimental data.

We then calculate the static conductivity tensor by using the quantum Kubo formula. The irreducible vertex part of the Bethe–Salpeter equation vanishes if evaluated consistently with the ansatz (3) for the self-energy. The components of the conductivity tensor (in units of e^2/h) become

$$\sigma_{\mu\mu} = \frac{2\pi}{r} (\hbar\omega_c)^2 \int dE \left(-\frac{df}{dE} \right) \sigma_{\mu\mu}(E) \quad (4)$$

with

$$\sigma_{\mu\mu}(E) = \sum_{n_1 n_2} \int d^2\theta |\langle n_1 \theta | k_\mu | n_2 \theta \rangle|^2 A_{n_1 \theta}(E) A_{n_2 \theta}(E) \quad (5)$$

for the longitudinal conductivity and

$$\sigma_{\mu\nu} = \frac{4}{r} (\hbar\omega_c)^2 \int dE f(E) \sigma_{\mu\nu}(E) \quad (6)$$

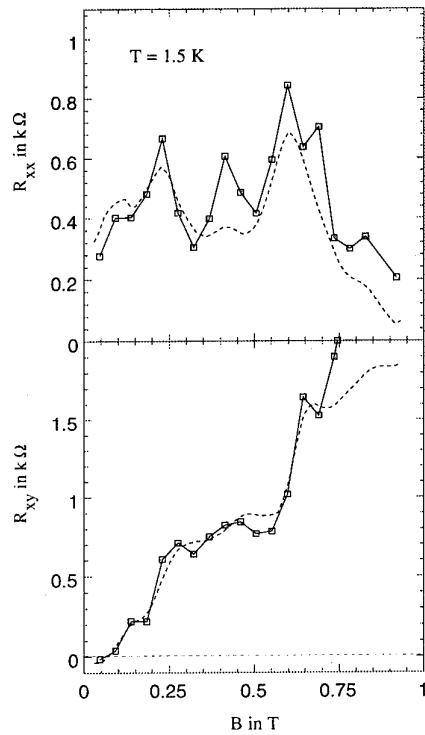


Figure 2. Longitudinal and Hall resistances for the potential of figure 1(a) (squares) in comparison with experimental data from [1] (broken curves). The full curves are guides for the eyes.

with

$$\sigma_{\mu\nu}(E) = \sum_{n_1 \neq n_2} \int \Im(\langle n_1 \theta | k_\mu | n_2 \theta \rangle \langle n_2 \theta | k_\nu | n_1 \theta \rangle) \times \Re \frac{dG_{n_1 \theta}}{dE} A_{n_2 \theta}(E) \quad (7)$$

for the Hall conductivity. \$A_{n\theta}\$ is the spectral function for the state \$|n, \theta\rangle\$ and \$\hbar\omega_c\$ is the cyclotron energy.

In contrast to weakly modulated 2D electron systems, for which the above formulae could be further simplified by treating the Landau level index as a good quantum number, Landau levels are strongly mixed in our case by the antidot potential, in particular for small magnetic fields. Thus, a completely numerical evaluation of both the longitudinal and Hall conductivities (from which we determine the resistances \$R_{xx}\$ and \$R_{xy}\$) has to be performed from equations (4)–(7) without simplifying approximations. Our results for \$R_{xx}\$ and \$R_{xy}\$ using the potential of figure 1(a) which models sample 3 of [1] are shown in figure 2 in comparison with the experimental data. All features of the experiment are reproduced by our quantum transport calculation: the commensurability peaks in \$R_{xx}\$, the non-quantized plateaus in \$R_{xy}\$, and the quenching of the Hall effect at very low magnetic fields. It is remarkable that we obtain quantitative agreement even on absolute scales for the resistance values. We also checked two additional experimental findings: a weak temperature dependence of the resistances and a shift of the commensurability peaks to higher magnetic fields with increasing electron density [1]. Both

properties are also clearly found in our calculated results (figure 3). A more detailed analysis of our calculations reveals the origin of these features to be correlations in the spectrum rather than the single-particle density of states [17]. This suggests a closer inspection of the distribution of the energy levels in our model system as a first step to clarify the nature of these correlations. By doing level statistics we should be able to detect features of the underlying classical phase-space structure in the quantum mechanical calculations.

3. Level statistics and quantum chaos

Level statistics with respect to quantum manifestations of classical chaos has to be performed in such a way that subspectra which are independent due to geometrical symmetries of the system are treated separately. The dominating geometrical symmetry is given here by the magnetic translation group, whose irreducible representations are labelled by the magnetic wavevector. Therefore level statistics is performed separately for each magnetic wavevector in the MBZ, which in our case leads to different statistics for different wavevectors. As we consider only fields where an integer number of flux quanta threads one unit cell of the lattice, the magnetic Brillouin zone equals the ordinary Brillouin zone. When proceeding to irrational magnetic field values, however, the MBZ shrinks to a point and the spacing of adjacent energy levels is determined by the width of the gaps. This has been done for a similar system in [15].

The calculated spectra become inhomogeneous on the energy scale, especially for larger magnetic fields, due to the clustering of the levels around Landau levels associated with unperturbed motion between the antidots. Therefore, we have to unfold the spectrum [6] before determining the level-spacing distribution.

There are three universality classes of the distribution of adjacent energy levels, which are associated with different ensembles of random matrix theory [6, 18]. They differ in the characteristic power \$\beta\$ of the level-spacing distribution \$P(s)\$ for \$s \rightarrow 0\$. It is \$\beta = 1\$ for the Gaussian orthogonal ensemble (GOE) corresponding to antiunitary symmetry of the Hamiltonian, \$\beta = 2\$ for the Gaussian unitary ensemble (GUE) for Hamiltonians without antiunitary symmetries, and \$\beta = 4\$ for the Gaussian symplectic ensemble (GSE) for time-reversal-invariant systems with Kramer's degeneracy but without geometrical symmetries respectively.

In order to characterize the level-spacing distribution in our case of an, in general, mixed classical phase space we use the phenomenological expression proposed in [12] for the intermediate case between the universality classes discussed above. It reads

$$P_\beta(s) = A \left(\frac{\pi s}{2} \right)^\beta \exp[-\frac{1}{16} \beta \pi^2 s^2 - (B - \frac{1}{4}\pi\beta)s] \quad (8)$$

and \$A\$ and \$B\$ are normalization constants. This distribution interpolates between Poisson statistics characterized by \$\beta = 0\$ and the GOE and GUE of random matrix theory.

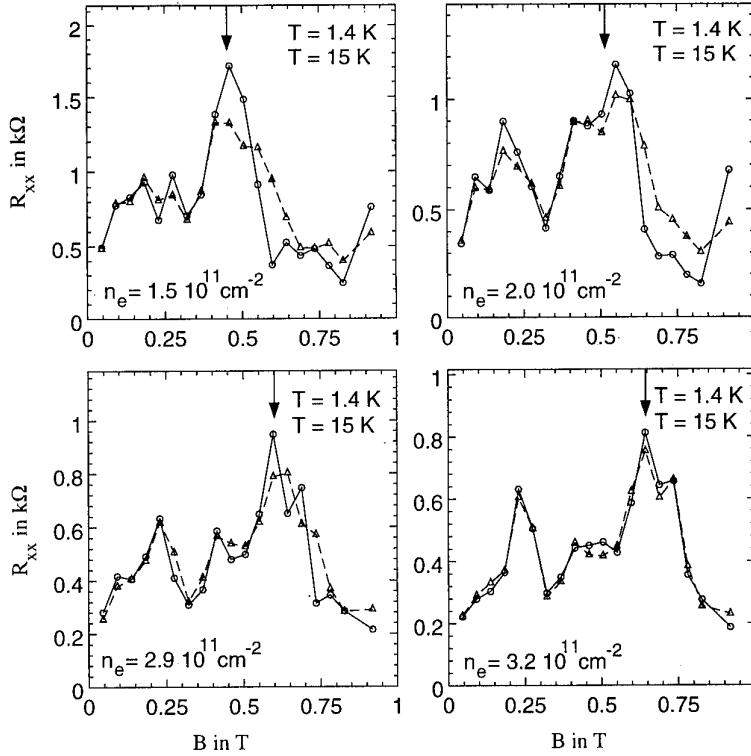


Figure 3. Longitudinal resistances for the potential of figure 1(a) for four different electron densities. Calculated values are given by circles (triangles) for the lower (higher) temperature and full curves are guides for the eye. The arrow indicates the position of the commensurate orbit around one antidot.

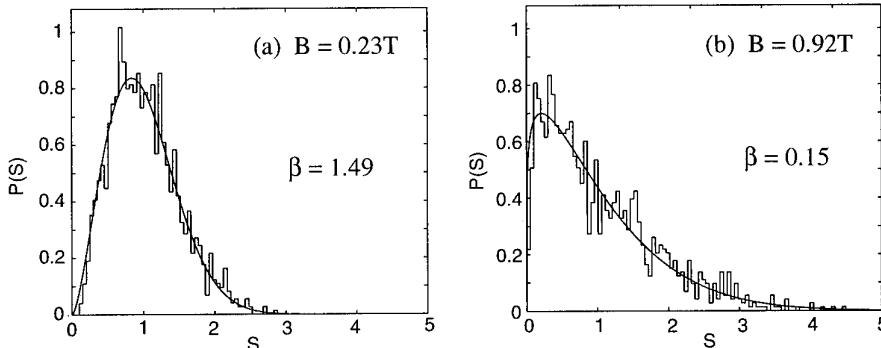


Figure 4. Distribution of nearest-neighbour spacings of energy levels for the potential ($V_0 = 25 \text{ meV}$, $\alpha = 4$ and $a = 300 \text{ nm}$) and the magnetic fields of figure 1. About 1500 levels have been included in the analysis. The magnetic wavevector has been chosen as $\theta = 0.77\pi, 0.33\pi$. The symbols are explained in the text.

For our model system the result for two different magnetic fields of $B = 0.23 \text{ T}$ and $B = 0.92 \text{ T}$ is shown in figure 4 together with a least-squares fit for the parameter β from equation (8). The value of the magnetic wavevector has been fixed at a point with no symmetries in the MBZ. For the lower magnetic field (figure 4(a)) the distribution is close to the one for the GUE statistics, whereas it shows the transition to a Poisson distribution on increasing the magnetic field to the higher value. This corresponds to the change of the classical phase-space

structure from a completely chaotic one for $B = 0.23 \text{ T}$ to one with large regular islands for $B = 0.92 \text{ T}$ as it is clearly demonstrated by the corresponding Poincaré sections. They are shown in figure 5 for an intermediate energy of 12.5 meV which is the Fermi energy for typical carrier densities in the experiment [1].

In the case of a completely chaotic phase-space structure (for low magnetic fields in our model) the universality class describing the fluctuations of level spacings and the corresponding ensemble of random

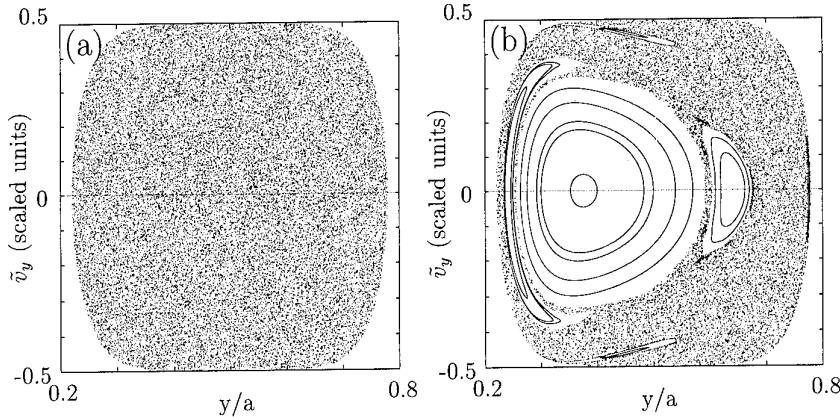


Figure 5. Poincaré sections for an intermediate energy of 12.5 meV. The potential is chosen as in figure 1(a), the magnetic fields are the same as in figure 4, i.e. $B = 0.23$ T (a) and $B = 0.92$ T (b).

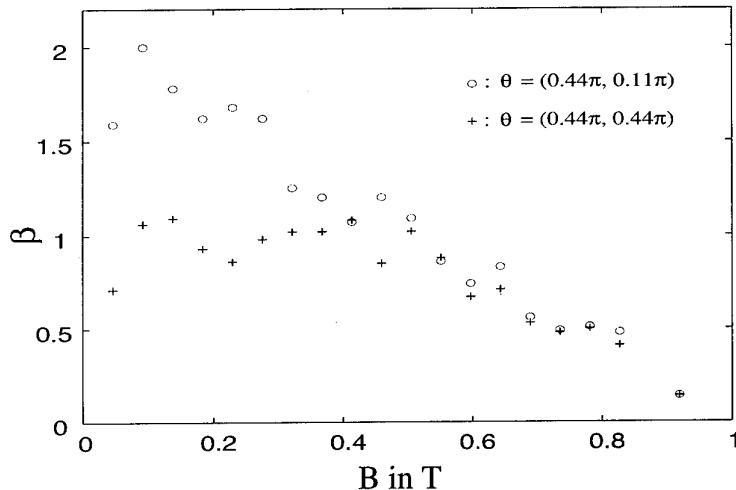


Figure 6. The parameter β of the level-spacing distribution equation (8) in dependence on the magnetic field for a point with no symmetries in the MBZ (open circles) and a point in the (1, 1) direction (crosses). The potential parameters are again chosen as in figure 1(a).

matrices depend on the space-time symmetries. Switching on an external magnetic field breaks the usual time-reversal symmetry [8]. For a spinless system a transition takes place from the GOE to the GUE, if there are no further geometric symmetries present in the system. In our model this should be observed for a general magnetic wavevector in the MBZ. The situation changes for wavevectors on symmetry lines as, for example, the (1, 1) direction. The combination of the operations of time reversal and reflection commutes with the Hamiltonian in that case and restores a generalized time-reversal symmetry [6]. The level statistics should then remain of the GOE type.

In figure 6 the parameter β is plotted against the magnetic field B for two different values of the magnetic wavevector θ corresponding to no symmetries (open circles) and to the (1, 1) direction (crosses) respectively. For low magnetic fields, one recognizes GUE statistics in

the former case, corresponding to $\beta = 2$, and GOE statistics ($\beta = 1$) in the latter case as discussed above. With increasing magnetic field the change of the classical phase-space structure corresponds to a transition to a Poisson distribution ($\beta = 0$). Note the form of the transition with the β coinciding down to a certain magnetic field and then changing to significantly different values in dependence on symmetry. For the lowest magnetic field value we find a decrease of β which can be ascribed to the fact that the symmetry-breaking effect of the magnetic field is not yet fully developed [8].

4. Conclusions

We have investigated quantum aspects of the carrier dynamics in antidot superlattices subject to a perpendicular magnetic field. Schrödinger's equation has been solved

for rational magnetic fields using the concept of magneto-translation group. Impurity scattering has been described in the self-consistent Born approximation and the Kubo formula has been used to obtain the longitudinal and Hall resistances. The calculated results reproduce quantitatively the experimental findings of structures in the magnetoresistance related to classical chaotic motion of the carriers. This includes commensurability peaks in the longitudinal resistance and non-quantized plateaus in the Hall resistance which is quenched for the lowest magnetic fields, as well as the dependence of amplitude and position of the commensurability peaks on temperature and electron density. To reveal quantum manifestations of the classical chaos in our system we have performed level statistics for points in the magnetic Brillouin zone with different symmetry. The results show a transition from level repulsion to level clustering with increasing magnetic field, corresponding to the transition from classically chaotic to regular motion of the charge carriers. Moreover, the universality class of level repulsion as described by random matrix theory has been found to depend on the symmetry properties in the magnetic Brillouin zone.

It is important to note that our model system is both available for experimental studies and shows different statistics of the energy levels as a function of the magnetic field as an external parameter. This is not easy to realize in complex atoms as the traditional experimental counterpart for theoretical investigations of quantum chaos. Thus, antidot superlattices should be an interesting object for future research in this direction.

To relate the quantum characteristics of a classically chaotic system to transport properties one has to investigate, in addition to the level statistics, the behaviour of the wavefunctions. It is known that these show localization in dependence on the character of the classical phase space [9]. In that way it should be possible to make contact with developments in the theory of disordered systems which yield universal properties of response functions. These ideas have been extended to chaotic systems by assuming that the average over the Fermi surface involved in the calculation of many

physical quantities is equivalent to the ensemble average for disordered systems [19].

Acknowledgments

We acknowledge financial support by the DFG (Sonderforschungsbereich 348, Teilprojekte A5 und YD).

References

- [1] Weiss D, Roukes M L, Menschig A, Grambow P, von Klitzing K and Weimann G 1991 *Phys. Rev. Lett.* **66** 2790
- [2] Schuster R, Ensslin K, Kotthaus J P, Holland M and Stanley C 1993 *Phys. Rev. B* **47** 6843
- [3] Yamashiro T, Takahara J, Takagaki Y, Gamo K, Namba S, Takaoka S and Murase K 1991 *Solid State Commun.* **79** 885
- [4] Fleischmann R, Geisel T and Ketzmerick R 1992 *Phys. Rev. Lett.* **68** 1367
- [5] Gerhards R R and Pfannkuche D 1992 *Phys. Rev. B* **46** 12606
- [6] Haake F 1991 *Quantum Signatures of Chaos* (Berlin: Springer)
- [7] Berry M V and Tabor M 1977 *Proc. R. Soc. A* **356** 375
- [8] Berry M V and Robnik M 1986 *J. Phys. A: Math. Gen.* **19** 649
- [9] Bohigas O, Tomsovic S and Ullmo D 1993 *Phys. Rep.* **223** 43
- [10] Brody T A 1973 *Lett. Nuovo Cimento* **7** 482
- [11] Berry M V and Robnik M 1984 *J. Phys. A: Math. Gen.* **17** 2413
- [12] Izrailev F M 1988 *Phys. Lett. A* **134** 13; 1989 *J. Phys. A: Math. Gen.* **22** 865
- [13] Lenz G and Haake F 1990 *Phys. Rev. Lett.* **65** 2325
- [14] Berry M V 1981 *Ann. Phys., NY* **131** 163
- [15] Geisel T, Ketzmerick R and Petschel G 1991 *Phys. Rev. Lett.* **66** 1651; 1991 *Phys. Rev. Lett.* **67** 3635
- [16] Silberbauer H 1992 *J. Phys.: Condens. Matter* **4** 7355
- [17] Silberbauer H and Rössler U 1994 *Phys. Rev. B* at press
- [18] Porter C E 1965 *Statistical Theory of Spectra: Fluctuations* (New York: Academic)
- [19] Taniguchi N and Altshuler B L 1993 *Phys. Rev. Lett.* **71** 4031

Resonant tunnelling quantum dots and wires: some recent problems and progress

P H Beton, H Buhmann, L Eaves, T J Foster, A K Geim,
N La Scala Jr, P C Main, L Mansouri, N Mori, J W Sakai and
J Wang

Department of Physics, University of Nottingham, Nottingham NG7 2RD, UK

Abstract. Electron tunnelling through donor-related states is discussed. This tunnelling process, which occurs well below the threshold voltage for conventional resonant tunnelling into the two-dimensional continuum states of the quantum well, reveals a new type of Fermi edge singularity effect which arises from the Coulomb interaction between the tunnelling electron on the localized site and the Fermi sea of electrons in the emitter layer. A new means of forming laterally confined resonant tunnelling devices is also described. By studying the effect of an applied magnetic field, the additional structure that appears in the current–voltage characteristics of these devices can be unambiguously associated with a lateral quantum mechanical confinement effect.

1. Introduction and overview

Double-barrier resonant tunnelling devices continue to attract considerable attention, even though it is now over 20 years since the original experimental work of Chang *et al* [1]. Those early experiments used large-area diodes in which the electrons move as free particles in the plane of the quantum well. In recent years, several groups have investigated the effect on the tunnel current versus voltage characteristics, $I(V)$, of confining laterally the electron motion in the plane of the well [2–8]. Such confinement produces additional quantization, which modifies the energy-independent density of states of an unconfined quantum well into a series of discrete peaks. Resonant tunnelling into the quantized states of a *one-dimensional quantum wire* or of a *zero-dimensional quantum box* should therefore give rise to additional peaks in $I(V)$.

In 1988, Reed *et al* [2] fabricated the first laterally confined resonant tunnelling devices, and observed additional resonant structure in $I(V)$, which they attributed to the lateral quantization effect. In this case, the lateral confinement arose from the physical shape of the structure, which was in the form of a *quantum pillar* with a confining potential provided by the surface charge and associated depletion charge. Devices of this form have been investigated further in recent work. In addition, three-terminal gated resonant tunnelling devices have also been successfully fabricated and studied [6, 7]. The effective cross-sectional area of these devices can be controllably decreased by applying a negative gate voltage to a metallic gate patterned in the form of a circular or square collar.

Due to their small cross-sectional area, laterally patterned resonant tunnelling devices have a small

capacitance and are therefore promising structures for the observation of single electron and Coulomb blockade phenomena at low biases close to the threshold for current flow [5]. However, electron tunnelling through a *single* unintentional donor impurity state in the quantum well of a small-area ($1 \mu\text{m}^2$) resonant tunnelling structure can give rise to an additional peak in $I(V)$ close to threshold [6]. Care has to be taken not to confuse this donor-related effect with lateral confinement and Coulomb blockade. The presence of donor impurities can arise not only from background contamination common to all growth processes, but also from impurity migration from the doped contact layers of the device [9]. The latter process can be significantly reduced by growing the layers at low temperatures and by increasing the width of the spacer layers between the tunnel barriers and the doped contact layers. By controlling donor segregation in this way, it has proved possible to introduce a specified concentration of donor impurities into the quantum well, as a low density δ -doped sheet [10]. A precise estimate of the spatial extent of the donor wavefunction has been obtained by investigating the effect of an applied magnetic field on the donor-assisted tunnelling peak in these structures.

This paper reviews some new developments involving the electrical properties of small-area resonant tunnelling diodes. The first of these is the observation of a Fermi edge singularity in resonant tunnelling between a two-dimensional electron gas (2DEG) and a strongly localized zero-dimensional state. We then describe a new route for the fabrication of small-area resonant tunnelling devices which relies on the use of wet etching and AlAs etch stops. The effect of a magnetic field on the tunnel current in these devices provides unequivocal evidence for lateral quantization.

2. A Fermi edge singularity in resonant tunnelling

Recently there has been considerable interest in the Fermi edge singularity (FES) effect in the optical spectra of doped semiconductors [11]. Electron-electron interactions may also influence the tunnel current [12], and theoretical work has suggested that tunnelling through a quantum dot may emerge as a useful tool for studying electron-electron interactions [13–18].

To investigate the 0D–2D tunnelling process we have employed our recent observation that the onset of the tunnel current in mesoscopic and also conventional, macroscopic resonant tunnelling devices (RTDs) is determined by tunnelling through random impurity-related states in the quantum well [10, 19]. Our technique is an alternative to the nanofabrication of quantum dots and provides much more strongly confined 0D states. The double-barrier RTDs were grown by molecular beam epitaxy (MBE) on n⁺GaAs substrates with substrate temperatures between 480 °C and 550 °C. The thickness of both (Al_{0.4}Ga_{0.6})As barriers is 5.7 nm, the quantum well width is 9 nm and there is a 20 nm spacer layer between each barrier and the more heavily n-doped contact regions. We also grew samples in which the centre plane of the quantum well was δ-doped by Si donors with concentrations of (2–8) × 10¹³ m⁻². Square mesas of side lengths varying between 6 μm and 100 μm were fabricated using photolithography.

Figure 1 shows a schematic energy band diagram for our devices under bias. A current flows when the energy of an electron in the emitter 2DEG is resonant with a state in the quantum well. The inset in figure 2 shows the main resonance due to the lowest 2D subband in the quantum well. At biases below the main resonance near the onset of the tunnel current we have found an additional step-like structure which is shown in figure 2 ($V > 70$ mV) for one of the undoped samples. Similar structure is seen in all devices, although details are unique to a particular device, and features occur in both directions of the applied bias but differ in their exact form. In the δ-doped samples the additional features are more numerous and extend

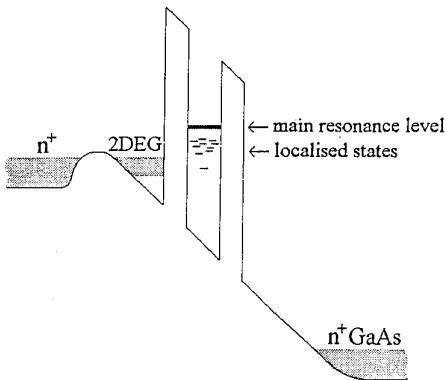


Figure 1. Schematic diagram of the conduction band profile of our devices under bias. Tunnelling occurs from a two-dimensional electron gas through the ground state in the quantum well (for the main resonance) or highly localized impurity levels at lower energies.

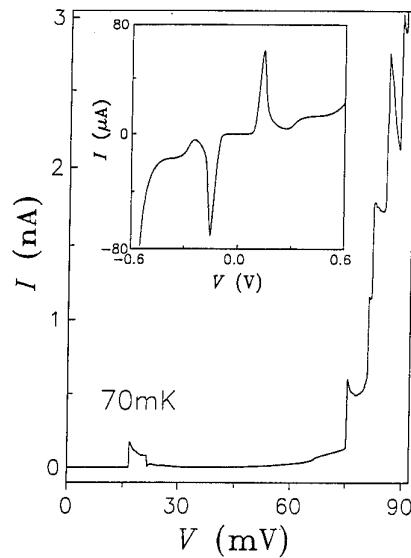


Figure 2. I(V) characteristics at low bias for a device 12 μm square. The inset shows the main resonance.

to lower values of applied bias. In many of the undoped devices isolated peaks occur such as that shown in figure 2 for $V \approx 20$ mV. We consider the additional features to be due to tunnelling through localized 0D states in the quantum well (qw) of the RTD with energies well below the edge of the lowest 2D subband as shown schematically in figure 1. The localized state giving rise to structure at 20 mV is also well below the expected binding energy of an isolated donor in the qw (~12 meV for a donor in the centre of the qw [10, 20]). We believe that a more tightly bound localized state of this kind is associated with a closely spaced donor pair, which is statistically likely in a mesa of this size (12 × 12 μm²).

Figure 3(a) shows in more detail the isolated peak of figure 2 and figure 3(b) is the current onset for a δ-doped RTD with 2×10^{13} m⁻² Si donors in the quantum well. The unexpected feature is the singular enhancement with decreasing temperature of tunnelling near the threshold (see figures 2 and 3) when the localized state is resonant with the emitter Fermi energy. Note that every current step in figure 2 is accompanied by such an enhancement. The characteristic width (full width at half-height) of the threshold peaks can be as small as 0.2 mV at the lowest temperature. The low-voltage edge of each step is thermally activated down to 70 mK. In general, as in figure 3(b), there is some additional oscillatory structure within the step at voltages above the threshold voltage V_{th} . However, in contrast to the singularity, this structure does not depend on temperature. The Fermi edge singularity is seen in *all* devices at temperatures below 1 K.

Impurity-assisted tunnelling, leaving aside the singularity, can be understood as follows. Under a typical applied bias of tens of mV, the tunnel current in our devices is limited by tunnelling through the emitter barrier and the states in the quantum well are empty

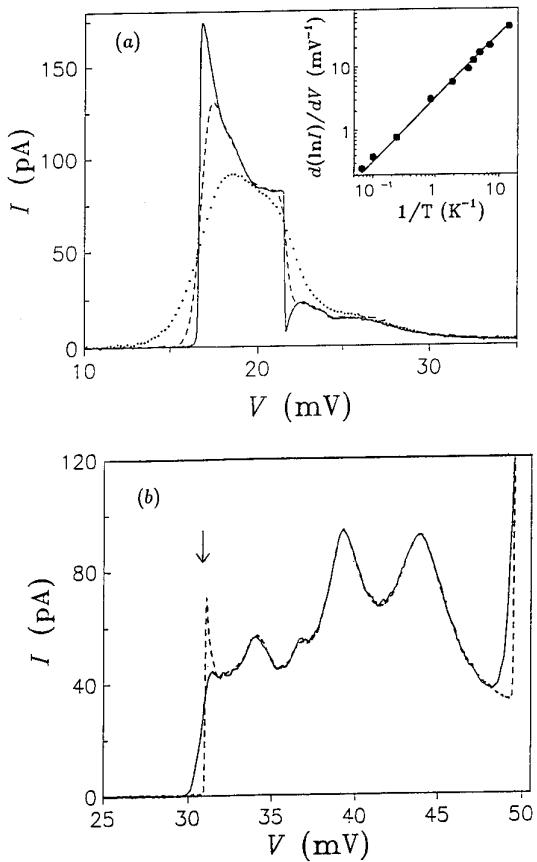


Figure 3. Detailed $I(V)$ characteristics at low biases when the first localized level is resonant with the emitter 2DEG. (a) The same device as in figure 2 at three different temperatures of 70 mK (full curve), 1.3 K (broken curve) and 5 K (dotted curve). (b) Another device 6 μm across and under opposite bias at 70 mK (broken curve) and 1.2 K (full curve). Inset: temperature dependence of the logarithmic slope of the tunnel current near the threshold voltage. The full line corresponds to $\alpha = 0.27$.

most of the time. As the bias increases, the impurity level moves downwards relative to the energy of the emitter 2DEG (see figure 1) and the tunnel current exhibits a step increase when the localized state coincides with the Fermi energy. As the voltage is increased further, and the energy of the 0D state becomes lower than the lowest energy state of the emitter, no states are available for resonant tunnelling and the current falls sharply (see figure 3(a)). In figure 3(b) a second impurity channel comes in resonance with the 2DEG (at $V \approx 50$ mV) before the first channel has passed away. The latter behaviour is also seen above 75 mV in figure 2. The tunnel current is given by

$$I = \frac{e}{\hbar} \Gamma_e(\varepsilon_i) \theta(\varepsilon_i) f(\varepsilon_i) \quad (1)$$

where $f(\varepsilon)$ is the Fermi distribution function, $\theta(\varepsilon)$ is the unit step function and ε_i is the energy of a 2DEG state resonant with the impurity state, measured from the bottom of the 2DEG subband. The 2D–0D tunnelling

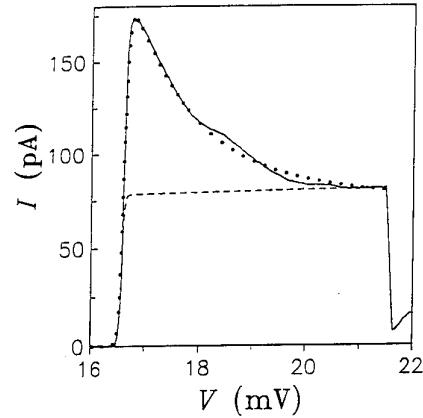


Figure 4. Comparison of the observed singularity with theory. The full curve is the experimental curve at 70 mK for the same device as in figures 2 and 3(a). The dotted curve is the behaviour expected from the ML theory. The broken curve: if the electron–electron interaction is neglected, the tunnel current within the step exhibits only a very small increase with increasing bias.

coefficient Γ_e can be written as

$$\Gamma_e(\varepsilon_i) = t \exp(-\varepsilon_i/\varepsilon_0) \quad (2)$$

where ε_0 is the binding energy of the localized state and t is a coefficient which includes parameters of the localized state and the tunnel barrier but is independent of the kinetic energy ε_i of tunnelling electrons within a 2DEG subband. Near the onset of the tunnel current, equation (1) varies as the Fermi function, which fits very well to the observed $I(V)$ characteristics. $I(V)$ curves for different temperatures intersect at the same point (for example see figure 3(a)), consistent with the case of 2D–0D tunnelling. To convert the voltage across the device into an energy scale we use $\varepsilon_F - \varepsilon_i = \alpha(V - V_{th})$, where α is characteristic of the distribution of electrostatic potential across the device. Experimental curves yield $\alpha = 0.25 \pm 0.05$ for all devices. The inset in figure 3(a) shows an example of the temperature dependence of the tunnel current below threshold at biases when $I \propto f(\varepsilon_i) \simeq \exp[(\varepsilon_F - \varepsilon_i)/k_B T]$ and hence $d \ln(I)/dV \simeq \alpha/k_B T$. The linear dependence of the logarithmic slope down to 70 mK indicates that the localized state has a very narrow linewidth. This is in agreement with the linewidth $\Gamma_e = \hbar/\tau_e \simeq 4$ mK expected from the tunnelling time $\tau_e = e/I \simeq 2$ ns.

Equation (2) shows that, for non-interacting electrons, the tunnel current within the step varies on the energy scale of the binding energy $\varepsilon_0 \sim 10$ –20 meV, much larger than typical values of the Fermi energy (1–5 meV) in the emitter accumulation layer for the first few steps in the $I(V)$ characteristics. Therefore, according to equations (1) and (2), variation of the current within the step is expected to be small, as shown by the broken curve in figure 4 calculated for the sample in figure 3(a). Clearly, the observed singularity in the tunnel current cannot be explained within a model involving only non-interacting electrons. Therefore, we attribute the FES to the influence of the electron–electron interaction. Its origin is the extra tunneling processes due to the Coulomb interaction

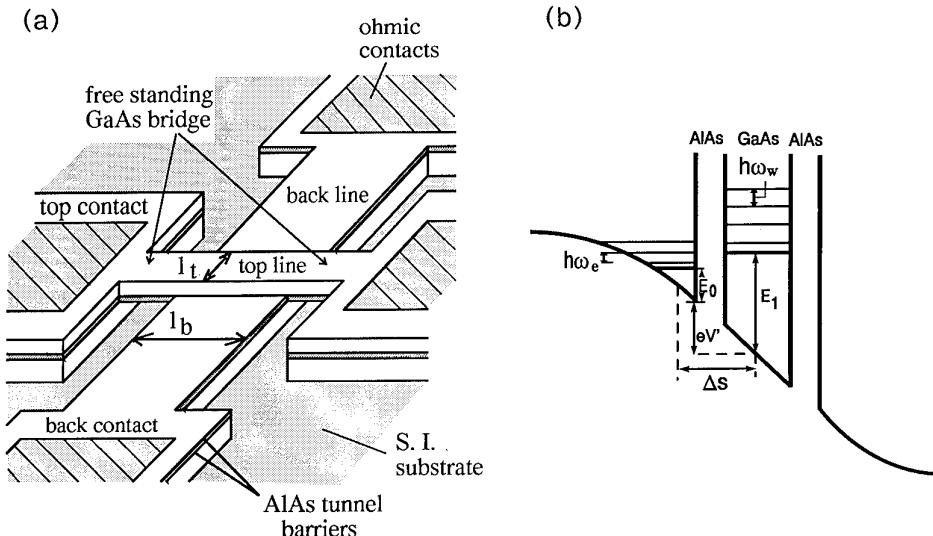


Figure 5. (a) A schematic diagram of our device. (b) Conduction band profile of our device.

between the fluctuating charge on the localized site and the Fermi sea in the emitter accumulation layer [17, 18]. The interaction allows an electron to violate the requirement of energy conservation between its initial and final states and the electron can tunnel into the localized state from an initial state which does not contribute in a model of non-interacting particles. The difference in the energies is transferred to or from the Fermi sea. A singularity arises at ϵ_F because scattering processes with small energy transfer are most effective (Fermi's golden rule) while the Pauli principle allows them only near ϵ_F . The theory of Matveev and Larkin [17] yields a power-law singularity of the form

$$I \propto (\epsilon_F - \epsilon_i)^{-\beta} \theta(\epsilon_F - \epsilon_i) \quad \beta \approx 3/8\pi^2(\lambda_F/d) \quad (3)$$

where λ_F is the Fermi wavelength. We estimate the distance from the plane of the 2DEG to the localized site (d) to be ≈ 25 nm, assuming the Fang–Howard approximation for the emitter 2DEG. Then, for the first few steps, which occur at biases between 15 and 80 meV, we expect β to be in the range 0.1–0.25. The interaction lasts for a finite time, τ_c , before an electron escapes from the impurity state into the collector contact. This leads to smearing of the singularity on the energy scale $\Gamma_c = \hbar/\tau_c$ as given by [17], so that

$$I \propto \left(\sqrt{(\epsilon_F - \epsilon_i)^2 + \Gamma_c^2} \right)^{-\beta} \left(\frac{\pi}{2} + \arctan \frac{\epsilon_F - \epsilon_i}{\Gamma_c} \right) \theta(\epsilon_F - \epsilon_i). \quad (4)$$

We estimate Γ_c in our devices to be of the order of 0.1 meV ($\tau_c \approx 10$ ps).

To describe the observed form of the FES we assume that the net current includes both the single-particle and many-body contributions given by equations (1) and (4) respectively. The absolute value of the many-body current in equations (3) and (4) is unknown and is used as a fitting parameter. Also, we allow Γ_c to vary around 0.1 meV to obtain the best agreement with the experimental data.

Figure 4 shows the best fit to the low-temperature $I(V)$ characteristic from figure 3(a). The coefficient β is calculated to be ≈ 0.22 for this bias and the fit yields $\Gamma_c \approx 0.2$ meV. For other samples, the singularities are also described by values of Γ_c within a factor of three of 0.1 meV. For completeness, to describe the observed temperature smearing at the onset of tunneling in figure 4 we have multiplied equation (4) by the Fermi function $f(\epsilon_i)$, instead of using the theta-function as in [17]. This allows us to fit the $I(V)$ curves for all temperatures below 1.5 K. For higher temperatures, when $k_B T > \Gamma_c$, the smearing of FES by temperature, in addition to Γ_c , becomes important.

3. Resonant magnetotunnelling via quantum confined states

This section describes a new processing technique to fabricate resonant tunnelling devices with submicrometre lateral dimensions. The active region of the device, grown by MBE on a semi-insulating substrate, consists of a GaAs quantum well (width $w = 9.0$ nm) formed between two AlAs tunnel barriers (thickness $b = 4.7$ nm). The growth temperature (550 °C) and layer parameters were chosen to minimize the incorporation of unintentional donors in the quantum well—in particular an undoped GaAs spacer layer (thickness $s = 20$ nm) was incorporated between each AlAs barrier and the adjacent doped GaAs contact layers. The active area of the device, shown schematically in figure 5(a), is the region of overlap of two GaAs bars, one (thickness l_t) etched in the top contact layer and the other (thickness l_b) in the lower contact layer. The AlAs barriers and GaAs quantum well are sandwiched in the region where the two bars overlap. The fabrication sequence, based on a series of selective wet etches, is described fully in [21]. A notable feature of the device is the use of a free-standing GaAs bridge to provide a connection to a large ohmic contact. Optical

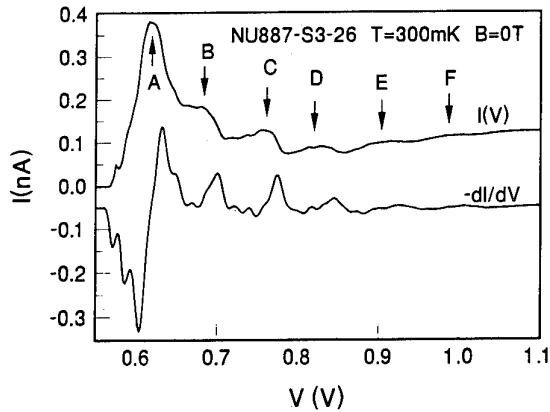


Figure 6. Low-temperature ($T = 0.3$ K) $I(V)$ and dI/dV for a small-area resonant tunnelling diode.

lithography and undercut etching are used to achieve submicrometre dimensions, l_t and l_b can be <0.5 μm , and typically $l_t < l_b$.

Figure 6 plots the $I(V)$ characteristics at 0.3 K for a device with $l_t = 0.5$ μm and $l_b = 1.0$ μm . In forward bias electrons flow to the top contact. From a comparison of diodes with different dimensions, we deduce that the sidewall depletion for these devices is ~ 0.2 μm for each edge and the current density is $J = 18$ nA μm^{-2} . The dimensions of the active conducting region are therefore ~ 100 nm \times 600 nm so that electrons tunnel between 1D quantum wires formed in the emitter and the quantum well. In the following discussion the x axis is the growth direction, and y and z are respectively perpendicular and parallel to the long axis of the wire.

Several clear peaks labelled A–F are observed in the $I(V)$ shown in figure 6. Peak A is the main resonance. The peak separation is ~ 70 mV. In addition to these peaks there is a much weaker series of peaks with a separation of ~ 20 mV. These peaks are resolved more clearly in the differential conductance (dI/dV) plot.

Figure 7 shows $I(V)$ for the same device in the presence of a magnetic field, B , oriented in the plane of the tunnel barriers, either perpendicular (figure 7(a)) or parallel (figure 7(b)) to the long axis of the quantum wire. We observe a clear anisotropy in the dependence of $I(V)$ on B for the two field orientations. For $B \parallel y$ (perpendicular to the wire, figure 7(a)), peaks A–F are shifted to higher voltage with increasing field. As B is increased beyond ~ 3 T the peaks become broader and eventually cannot be resolved. In addition for $B > 4$ T a weak series of features develops in the rise to the main resonance, identified by arrows in figure 7(a) for the $B = 10$ T curve. For $B \parallel z$ (parallel to the wire, figure 7(b)), peaks A–F show a complex splitting for small magnetic fields. However, for $B > 3$ T, a regular series of peaks are observed in $I(V)$ with a typical spacing of ~ 30 mV. These can be seen more clearly in the differential conductance plots in figure 8. Remarkably we are able to observe more than 15 peaks in this series, as labelled in figure 7(b).

To explain our data we have developed a simple model for resonant tunnelling through laterally bound

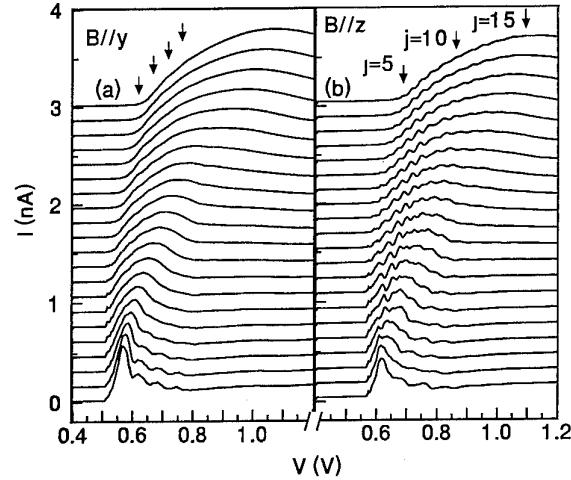


Figure 7. $I(V)$ at $T = 0.3$ K in the presence of a magnetic field between $B = 0$ T (lowest curve) and $B = 10$ T (top curve) in 0.5 T steps. The field is oriented perpendicular to the current flow and either perpendicular (a) or parallel (b) to the axis of the quantum wire.

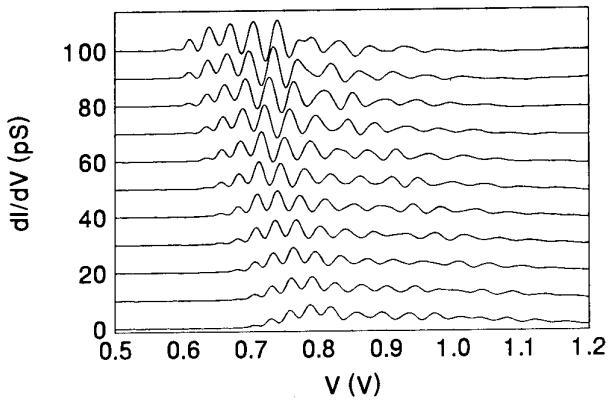


Figure 8. Plot of differential conductance dI/dV versus V for B parallel to the axis of the quantum wire. Bottom curve is at $B = 12$ T with 0.5 T intervals to top curve at $B = 7$ T.

states. For a large-area diode the confinement energy of the lowest quasibound state in the quantum well is E_1 (~ 40 meV for our heterostructure), and the bound state energy of the two-dimensional accumulation layer formed under bias at the emitter/barrier interface is E_0 (see figure 5(b)). For our small-area diode we assume that the well and emitter states are laterally quantized in the y direction by a parabolic potential with energy levels spaced by $\hbar\omega_w$ and $\hbar\omega_e$ respectively. From our previous work [22] on small-area RTDS we expect the degree of confinement to be different in the emitter and the well, and that in forward bias $\hbar\omega_w > \hbar\omega_e$. Free electron motion is assumed for the z direction. Thus a set of 1D subbands are formed in the emitter (denoted by index i) and well (index j).

The tunnel current can be calculated using the transfer Hamiltonian formalism. Full details can be found in [21] (see also [23–25] for similar treatments of low dimensional tunnelling problems).

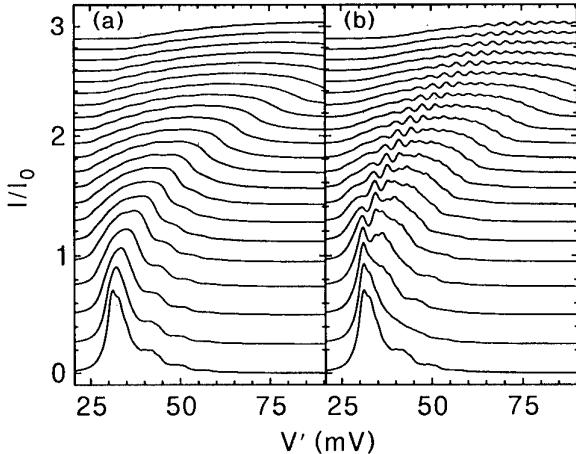


Figure 9. Calculated plots of I/V for comparison with experimental data in figure 7. The calculated curves are for fields between $B = 0$ T (lowest curve) up to $B = 10$ T (top curve) in 0.5 T steps. Figure 9(a) ((b)) is for a magnetic field oriented perpendicular (parallel) to the axis of the quantum wire. Parameters are discussed in the text.

Figures 9(a) and 9(b) show the calculated current as a function of V' (V' is the voltage dropped between the emitter/tunnel barrier interface and the centre of the quantum well—see figure 5(b)) in the presence of a magnetic field perpendicular (figure 9(a)) or parallel (figure 9(b)) to the wire axis. The following parameters were used: $\hbar\omega_w = 3.5$ meV, $\hbar\omega_e = 1.5$ meV, $E_F = 6$ meV, $\Gamma = 1.5$ meV. The value of E_F , the Fermi energy in the emitter, is extracted from magneto-oscillations in the tunnel current of a large-area RTD. Γ is a parameter characterizing the width of the energy levels. The choice of the other parameters is justified below. There is remarkable agreement between the calculated and experimental curves. In particular the calculated curves show weak and strong peaks with approximately the correct relative magnitudes with a magnetic dependence that corresponds closely to experiment for each field orientation.

The effect of field for the perpendicular case, $B \parallel y$, is similar to that of a large-area diode [26] for which the voltage threshold for resonant tunnelling shifts to a lower value for low field ($k_0 < k_F$, where $k_0 = eB\Delta s/\hbar$, and Δs is the spatial separation in the x direction of the emitter and well state), followed by a progressive increase in the threshold voltage and a broadening of the resonance as B is increased ($k_0 > k_F$). For our small-area diode there are several occupied 1D subbands in the emitter, each with a different Fermi wavevector

$$k_F^i = 2m[E_F - (i + \frac{1}{2})\hbar\omega_e]^{1/2}/\hbar.$$

Each subband therefore has a different voltage threshold for resonant tunnelling. This accounts for the weak structure we observe in our data at high field (see arrows in figure 7(a)) which is reproduced in our calculated curves. We deduce that there are four 1D subbands below the Fermi energy, $E_F/\hbar\omega_e \sim 4$, corresponding to the value used in our calculations. The additional peaks B, C . . .

together with the weaker series at $B = 0$ T are entirely due to the variation in coupling (i.e. the overlap integral) between 1D subbands as a function of applied voltage. This coupling is unaffected for $B \parallel y$ so each of the additional peaks is expected to have approximately the same dependence on B as the main resonance in accordance with our experimental data.

For parallel field, $B \parallel z$, the coupling between 1D states is strongly modified. We consider the limit $E_F/\hbar\omega_e \rightarrow \infty$ (i.e. taking the emitter states as plane waves), which is valid if $\hbar\omega_e \ll \hbar\omega_w$ and $\hbar\omega_e \ll E_F$. In this limit the current is given by

$$I = I_0 \sum_j \int_{-k_F+k_0}^{k_F+k_0} dk_y |\Phi_j(k_y)|^2 \frac{\Gamma}{\Gamma^2 + \Delta E_j^2(V', B)} F(k_y) \quad (5)$$

where

$$\begin{aligned} \Delta E_j = E_0 - E_l - (j + \frac{1}{2})\hbar\omega_w + (\alpha_e - \alpha_w)B^2 \\ + eV' + \hbar^2 k_y^2/2m, \end{aligned}$$

k_y is the wavevector of the emitter state, I_0 is a constant and Φ_j is the Fourier transform of the wavefunction of the j th harmonic state. $F(k_y) = [1 - (k_y - k_0)^2/k_F^2]^{1/2}$ is a term which arises from integrating over k_z . The B^2 term in ΔE_j is a diamagnetic shift, and α_e and α_w are determined by the spatial extent of the wavefunction in the x direction for the emitter and well respectively. Previously we have found [10] that $\alpha_e > \alpha_w$ for similar MBE layers. Thus tunnelling via a state j occurs over a voltage range for which $\Delta E_j < \Gamma$. For given j and k_y , $\Delta E_j = \text{constant} + (\alpha_e - \alpha_w)B^2 + eV'$, so that the voltage required for tunnelling via state j is reduced as B is increased. This accounts for the shift to lower voltage for the high-field peaks in both our experimental data and our calculated curves.

The tunnel current flowing from state k_y via state j is determined by the overlap between initial and final states, $\Phi_j(k_y)$. Figure 10 shows $|\Phi_j|^2$ up to $j = 15$. The high-energy states have pronounced maxima in $|\Phi_j|^2$ at a value $k_{j,\max} = (\pm [2m(j + \frac{1}{2})\hbar\omega_w]^{1/2}/\hbar)$. As B is increased, these maxima are the only features in $|\Phi_j|^2$ which remain within the integration range (centred around $k_y = k_0$). This is true, for example, for states $j = 6-11$ at $B = 10$ T (see figure 10). At this value of B tunnelling via each of the states $j = 6-11$ therefore results in a clear isolated peak in $I(V)$. A further increase in B moves the integration range beyond $k_{j,\max}$ and the tunnel current via the j th state falls to zero (this is true for $j = 1-5$ at $B = 10$ T; see figure 10). This corresponds to the high-field amplitude dependence of our observed peaks. Furthermore, at high field there is only one peak in $I(V)$ due to tunnelling via a particular laterally quantized state, as implied by the labelling $j = 1-15$ in figure 7(b).

For $B = 0$, peaks in $I(V)$ occur when the condition $\Delta E_j(k_y) < \Gamma$ is met, so generally a given state will give rise to n peaks in $I(V)$ where n is the number of maxima in $|\Phi_j|^2$ lying in the range $0 < k_y < k_F$. These peaks

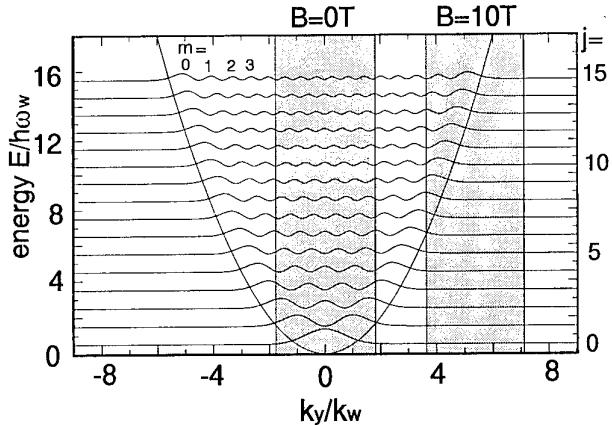


Figure 10. Plots of probability density $|\Phi_j|^2$ versus k_y/k_w for the laterally bound states in the quantum well. Plots are shown for the lowest 15 states. Also shown (shaded) is the integration range $k_0 - k_F < k_y < k_0 + k_F$ for equation (5) for $B = 0$ T and $B = 10$ T. The horizontal axis is in units $k_w = (m\omega_w/\hbar)^{1/2}$. From our data we deduce $\hbar\omega_w = 3.5$ meV, $E_F = 6$ meV and $\Delta s = 20$ nm. These values were used to calculate the curves in figure 8. This gives $k_w = 5.6 \times 10^7$ m $^{-1}$, $k_F = 1.0 \times 10^8$ m $^{-1}$ and k_0 ($B = 10$ T) = 3.0×10^8 m $^{-1}$.

correspond to the weak structure observed in both our data and our calculated curves. The series of stronger peaks (A–F) are due to a modulation in the matrix element. Full details of how these effects give rise to the form of the $I(V)$ curves in zero magnetic field are given in [21].

4. Conclusion

In this paper, we have shown that low-temperature current–voltage characteristics of small-area resonant tunnelling diodes continue to yield new physical phenomena. We have described how resonant tunnelling through donor-related states in the quantum well reveals a new Fermi edge singularity effect in quantum transport. We have also described a new fabrication technique for producing resonant tunnelling devices with lateral confinement. By using high magnetic fields, we have shown unambiguously that this confinement effect gives rise to laterally quantized states and associated structure in the $I(V)$ characteristics. The open structure of these devices also opens up the prospect of using a p-i-n variant as a resonant tunnelling light emitting quantum dot [27].

Acknowledgments

We are grateful to Dr M Henini for growing the layers used in this work. This work is supported by the Science and Engineering Research Council (UK) and also, in part, by the EC Basic Research Action no 7193 (PARTNERS). PHB and LE are grateful to the Royal Society for financial support.

References

- [1] Chang L L, Esaki L and Tsu R 1974 *Appl. Phys. Lett.* **24** 593
- [2] Reed M A, Randall J N, Aggarwal R J, Matyi R J, Moore T M and Wetsel A E 1988 *Phys. Rev. Lett.* **60** 535
- [3] Tewordt M, Law V J, Kelly M J, Newbury R, Pepper M, Peacock D C and Frost J E F 1990 *J. Phys.: Condens. Matter* **2** 8969
- [4] Tarucha S, Hiarayama Y, Saku T and Kimura T 1990 *Phys. Rev. B* **41** 5459
- [5] Su B, Goldman V J, Santos M and Cunningham J E 1991 *Appl. Phys. Lett.* **58** 747
- [6] Dellow M W, Beton P H, Langerak C J G M, Foster T J, Main P C, Eaves L, Henini M, Beaumont S P and Wilkinson C D W 1992 *Phys. Rev. Lett.* **68** 1754
- [7] Guérat P, Blanc N, Germann R and Rothuizen H 1992 *Phys. Rev. Lett.* **68** 1896
- [8] Ramdane A, Faini G and Launois H 1991 *Z. Phys. B* **85** 389
- [9] Harris J J, Clegg J B, Beall R B, Castagne J, Woodbridge K and Roberts C 1991 *J. Crystal Growth* **111** 239
- [10] Sakai J-W, Fromhold T M, Beton P H, Eaves L, Henini M, Main P C, Sheard F W and Hill G 1993 *Phys. Rev. B* **48** 5664
- [11] Skolnick M S, Rorison J M, Nash K J, Mowbray D J, Tapster P R, Tapster S J, Bass S J and Pitt A D 1987 *Phys. Rev. Lett.* **58** 2130
Hawrylak P 1992 *Phys. Rev. B* **45** 8464 and references therein
- [12] For a review see Altshuler B L and Aronov A G 1985 *Electron–Electron Interactions in Disordered Systems* ed A L Efros and M Pollak (Amsterdam: Elsevier Science) ch 3, p 27
- [13] Glazman L I and Matveev K A 1988 *Pis'ma Zh. Eksp. Teor. Fiz.* **48** 403 (Engl. transl. 1988 *JETP Lett.* **48** 445)
- [14] Herschfield S, Davies J H and Wilkins J W 1991 *Phys. Rev. Lett.* **67** 3720
- [15] Yeyati A L, Martin-Rodero A and Flores F 1993 *Phys. Rev. Lett.* **71** 2991
- [16] Ralph D C and Buhrman R A 1992 *Phys. Rev. Lett.* **69** 2118
- [17] Matveev K A and Larkin A I 1992 *Phys. Rev. B* **46** 15337
- [18] Bauer G E W 1994 *Surf. Sci.* **305** 358
- [19] Sakai J W et al 1994 *Appl. Phys. Lett.* **64** 2563
- [20] Greene R L and Bajaj K K 1983 *Solid State Commun.* **45** 825
- [21] Wang J, Beton P H, Mori N, Eaves L, Buhmann H, Mansouri L, Main P C, Foster T J and Henini M 1994 *Appl. Phys. Lett.* at press
- [22] Beton P H, Dellow M W, Main P C, Eaves L, Henini M, Foster T J, Beaumont S P and Wilkinson C D W 1992 *Appl. Phys. Lett.* **60** 2508
- [23] Demmerle W, Smoliner J, Gornik E, Bohm G and Weimann G 1993 *Phys. Rev. B* **47** 13547
- [24] Snell B R, Chan K S, Sheard F W, Eaves L, Toombs G A, Maude D K, Portal J C, Bass S J, Claxton P, Hill G and Pate M A 1987 *Phys. Rev. Lett.* **59** 2806
- [25] Bryant G W 1989 *Phys. Rev. B* **39** 3145
- [26] Leadbeater M L, Eaves L, Simmonds P E, Toombs G A, Sheard F W, Claxton P A, Hill G and Pate M A 1988 *Solid-State Electron.* **31** 707
see also Smoliner J, Demmerle W, Berthold G, Gornik E, Weimann G and Schlapp W 1989 *Phys. Rev. Lett.* **63** 2116
- [27] Buhmann H, Mansouri L, Wang J, Beton P H, Eaves L and Henini M 1994 *Preprint*
see also Buhmann H, Wang J, Mansouri L, Beton P H, Eaves L, Heath M and Henini M 1994 *Solid-State Electron.* **37** 973

Low-dimensional resonant tunnelling and Coulomb blockade: a comparison of fabricated versus impurity confinement

M R Deshpande[†], E S Hornbeck[‡], P Kozodoy[†], N H Dekker^{†§},
J W Sleight^{||}, M A Reed^{‡||}, C L Fernando[¶] and W R Frensley[¶]

[†] Department of Physics, Yale University, New Haven, CT 06520, USA

[‡] Department of Electrical Engineering, Yale University, New Haven, CT 06520, USA

^{||} Department of Applied Physics, Yale University, New Haven, CT 06520, USA

[¶] Department of Electrical Engineering, Erik Jonsson School of Engineering,
University of Texas at Dallas, Richardson, TX 75083, USA

Abstract. Nanometre-scale fabrication techniques, combined with epitaxial resonant tunnelling structures, now routinely allow the study of quasi-0D confined electron systems. In addition to energy level separations that are tunable by the confining potentials, these systems can also exhibit Coulomb blockade. Surprisingly similar effects are also observed for conventional, unconfined resonant tunnelling devices. We have recently discovered that the turn-on characteristics of nearly all resonant tunnelling devices exhibit sharp peaks in conductance, attributable to tunnelling through single quantum well donor states. These unintentional donor states are distributed in energy, depending on position in the quantum well. We have performed electronic spectroscopy of these states, and consistently find binding energies approximately 10 meV greater than expected for a single quantum well donor due to quantum well fluctuations. We present detailed measurements of single-electron tunnelling through a *single* donor bound state utilizing simple non-confined heterostructures.

1. Coulomb blockade and quantum dot states in laterally confined vertical resonant tunnelling heterostructures

Advances in nanofabrication have recently permitted the investigation of low-dimensional states in laterally confined heterostructures. A strategy to probe the low-dimensional states is vertical tunnelling, of which one embodiment is a laterally confined vertical resonant tunnelling heterostructure. Transport in these structures [1–11] exhibits a rich variety of effects, such as tunnelling via quasi-zero-dimensional (“quantum dot”) states and Coulomb blockade. It has recently become apparent that the epitaxial and fabrication details of the device are crucial in determining which effects will be dominant. In the simple cases of a confined vertical resonant tunnelling structure, quasi-zero-dimensional states are observed (neglecting the possibility of tunnelling through impurity states, which will be discussed later). In addition, Coulomb blockade will be present when the collector barrier is approximately equal to, or greater than, the emitter barrier [12]. This allows the simultaneous observation of Coulomb blockade (CB) and quasi-zero-dimensional states, as has been experimentally realized [9, 10] by a judicious choice of tailored barrier heights

and relatively large lateral size. However, in the interesting case of strong lateral confinement (for example of the order of tens of meV laterally confined state splittings, which are superimposed onto the vertical heterostructure confinement energy E_{QW}), it becomes difficult in the GaAs/AlGaAs system to access the Coulomb blockade regime. This is because the depletion layers arising from the lateral confinement superimpose upon the internal contact potential between the doped contacts and the undoped resonant tunnelling region, resulting in an exaggerated internal contact potential [2]. The additional potential now required to bring the quantum well states into resonance often results in substantial collector barrier lowering, which is why some experiments exhibit only quasi-zero-dimensional states.

A way to access both large lateral confinement and Coulomb blockade is to reduce the voltage required to bring the dot states (i.e. those arising from the $n = 1$ quantum well state of the unconfined structure) into resonance. Vertically enlarging the thickness of the quantum well brings the additional complication of reducing $E_{QW,n=2} - E_{QW,n=1}$, and analysis becomes difficult due to overlap of CB and dot states. An alternative, realized here, is to ‘reset’ the zero of the $E_{QW,n=1}$ state by using an InGaAs quantum well instead of GaAs. Figure 1 shows the conduction band structure of the (unconfined) epitaxial design used here, modelled fully self-consistently using nominal growth parameters.

[¶] Present address: Physics Department, Leiden University, The Netherlands.

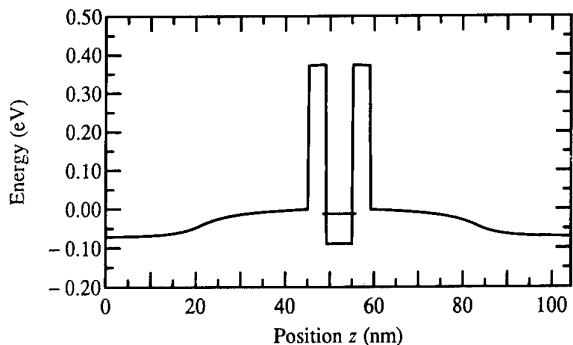


Figure 1. Self-consistent conduction band diagram, using nominal growth parameters, of the GaAs contact/AlGaAs barrier/InGaAs quantum well tunnelling structure used to fabricate quantum dots.

Figure 2 shows the $I(V)$ characteristics at 1.4 K of a quantum dot of approximately 150 nm diameter using a GaAs contact ($n^+ \sim 1.5 \times 10^{18} \text{ cm}^{-3}$ Si)/Al_{0.3}Ga_{0.7}As barrier/In_xGa_{1-x}As ($x \sim 0.08\text{--}0.1$) quantum well/Al_{0.3}Ga_{0.7}As barrier/GaAs contact vertical structure. The fabrication of these types of structures has been previously reported [1]. There are two main features evident: characteristic steps (plateaus) in the $I(V)$ at low bias with a few mV separation, and a weak structure superimposed onto the familiar NDR characteristic at higher bias (near and above the NDR peak) with a characteristic splitting of ~ 20 mV. Variable-temperature measurements up to 70 K indicate distinct behaviour between the low- and high-bias structures; the low-bias structure disappears by 10 K, whereas the higher-bias structure is still evident at $\sim 40\text{--}50$ K. The characteristic shape and temperature dependence of the higher-bias structure is very similar to other results [1–11], indicative of 0D (zero-dimensional) quantum dot states of ~ 10 meV splittings.

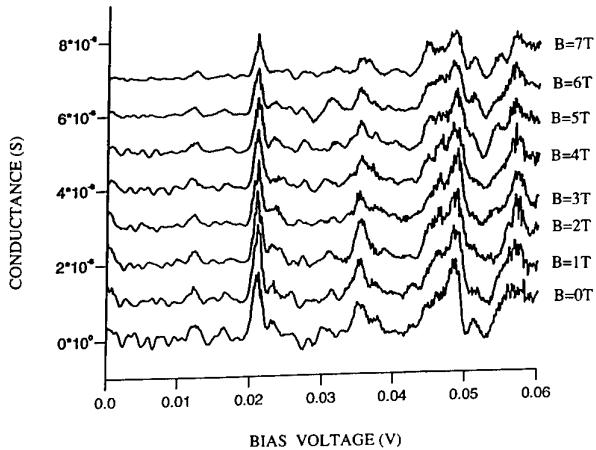


Figure 3. $G(V)$ versus B (0 to 7 T in 1 T increments, increasing upward) characteristics at 1.4 K of the quantum dot of figure 1.

The low-bias structure is suggestive of Coulomb blockade [9, 10]. To determine the charging energy, the voltage separation of the first and second plateaus is measured to be 3 mV, resulting in a capacitance of $2.7 \times 10^{-17} \text{ F}$, assuming vertical symmetry. In a simple cylindrical parallel-plate approximation, this gives a diameter of 36 nm, in good agreement with the resultant conductive ‘core’ after subtracting the expected lateral depletion produced by the free surfaces. As an additional verification, figure 3 shows $G(V)$ versus B measurements. Whereas the quantum dot states are sensitive to magnetic field [11], the Coulomb blockade states are not sensitive to B , as observed. Finally, it is noted that the low-bias Coulomb blockade structure weakens with increasing bias, disappearing by ~ 0.1 V, as would be expected with significant collector barrier lowering.

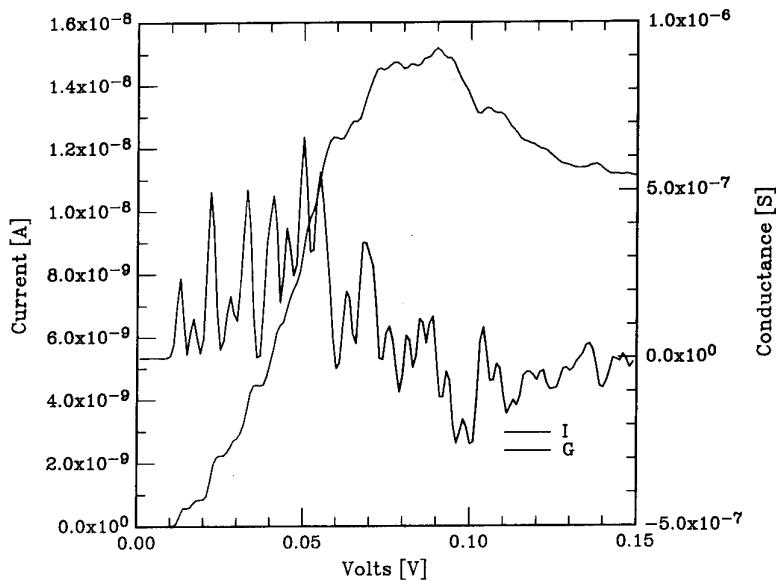


Figure 2. $I(V)$ and $G(V)$ characteristics at 1.4 K of a quantum dot of approximately 150 nm diameter using a GaAs contact ($n^+ \sim 1.5 \times 10^{18} \text{ cm}^{-3}$ Si)/Al_{0.3}Ga_{0.7}As barrier/In_xGa_{1-x}As ($x \sim 0.08\text{--}0.1$) quantum well/Al_{0.3}Ga_{0.7}As barrier/GaAs contact vertical structure.

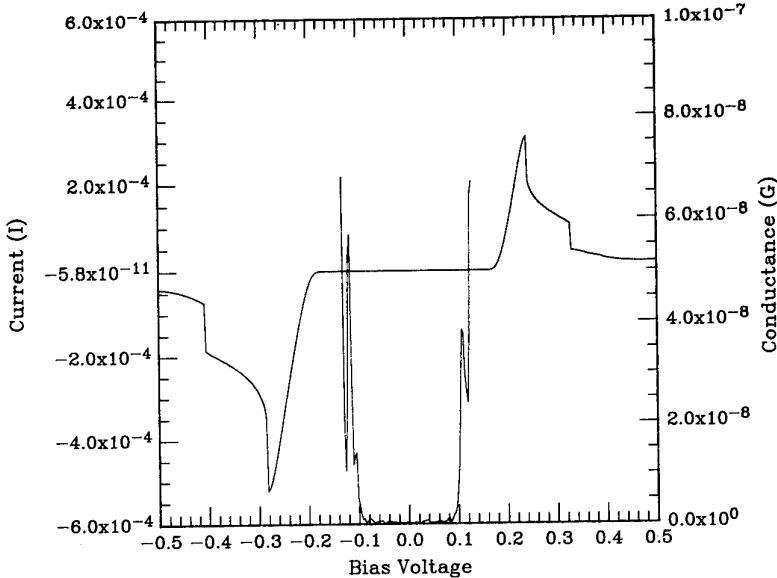


Figure 4. $I(V)$ and magnified $G(V)$ (in the turn-on region) of a large-area resonant tunnelling diode at 4.2 K. The epitaxial structure is nominally a 4.5 nm GaAs quantum well with 8.5 nm AlGaAs barriers.

2. Tunnelling through donor bound states

The characteristic plateau structure of Coulomb blockade (see previous section) has been reported for submicrometre resonant tunnelling structures (see, for example [7]; or in the turn-on region of numerous other experiments) but is recognized as being quantitatively inconsistent with both Coulomb blockade and energy quantization. Recently, this type of structure in sub-micrometre structures has been attributed to tunnelling through donor bound states in the quantum well [13–15]. Analysis of the conductance values, and a comparison of the experimental spectrum with the estimated confinement and Coulomb blockade versus donor energies, are suggestive that these are indeed donor bound states, though the possibility of lateral confinement effects in these structures complicates the analysis. Observation of the donor states should be possible even for larger (unconfined) lateral sizes, giving an unambiguous interpretation [16]. Measurements similar to the following have been reported [14], but for an intentionally δ -doped quantum well. Here we present results for an undoped well, giving a sufficiently dilute (unintentional) donor concentration such that the regime of tunnelling through a single impurity can be accessed.

We have observed that the states due to tunnelling through a single impurity also exist for non-confined resonant tunnelling structures; indeed, we have observed such structure for every resonant tunnelling structure we have examined (for which sufficient current resolution is possible). Figure 4 shows the $I(V)$ and magnified $G(V)$ (in the turn-on region) of a large-area resonant tunnelling diode. The epitaxial structures used in this study are identical to the highly characterized structures previously reported [17]; briefly, a set of three structures with 11.8 nm, 8.5 nm or 6.5 nm AlGaAs barriers cladding

a nominally 4.4 nm GaAs quantum well. Figure 4 presents results from the 8.5 nm barrier sample. Figure 5 shows in detail the turn-on region. Note that the fine structure of interest is more than six orders of magnitude smaller in current than the main resonance. The current has well defined steps (plateaus) with conductance widths (FWHM) of ~ 1.5 mV at 1.4 K, which are nearly symmetric in voltage position, magnitude and width around zero bias. Numerous devices were examined, all showing similar structure. The number of peaks observed approximately scaled with the area of the device (from $2 \mu\text{m}^2$ to $64 \mu\text{m}^2$) to the extent that distinct peaks were observable, both for the 8.5 nm and 6.5 nm barrier devices. (A similar study for the 11.8 nm structure was not possible due to minimum current detection limits.) The conductance maxima also decreased appropriately with increasing barrier thickness, a factor of about ten between the 6.5 nm and 8.5 nm devices.

If these are unintentional dopants, an estimate of the number of impurities in the quantum well can be made; for a $\sim 1 \times 10^{14} \text{ cm}^{-3}$ background, we should expect about two in the quantum well of a $2 \mu\text{m}^2$ device. Indeed, only two distinct peaks are observed throughout the turn-on region of a $2 \mu\text{m}^2$ device (data not shown here). For larger devices (such as figures 4 and 5), lower-voltage peaks are isolated and distinct, though it becomes more difficult to determine the total number of peaks, due to overlap. However, as noted, the number of apparent peaks increases appropriately with increasing device size. Comparing these results to those of [14], the present results allow one to examine in a simple unconfined tunnelling structure single-electron tunnelling through a *single* impurity (analogous to lateral MOSFET experiments [18]) versus a saturated ensemble.

The observed current due to tunnelling through donor bound quantum well states should be related only

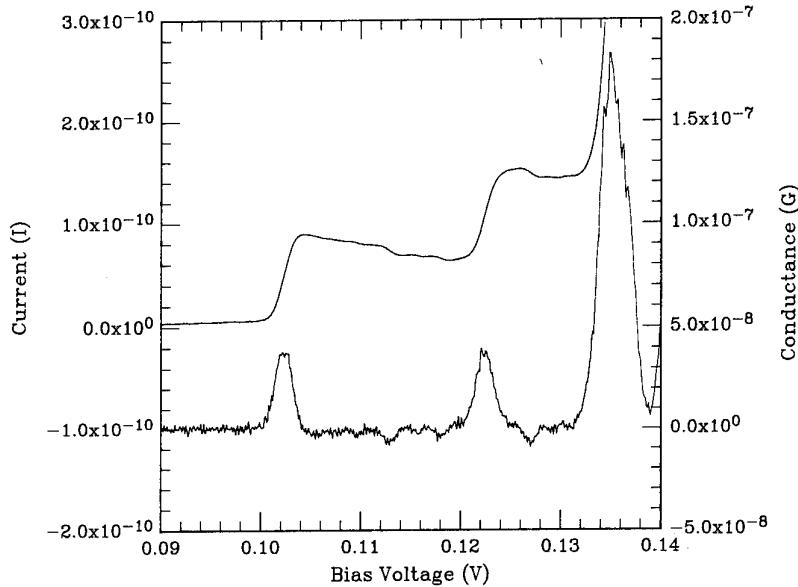


Figure 5. $I(V)$ and $G(V)$ characteristics of an $8 \mu\text{m}^2$ device for the 8.5 nm barrier sample, at 1.4 K.

to the lifetime of the state, in the low-bias and low-temperature limit, by $\Delta I = e/\tau$. Figure 5 illustrates a constant ΔI of $\sim 8 \times 10^{-11} \text{ A}$, giving a lifetime of $2 \times 10^{-9} \text{ s}$. An increase by about a factor of ten for the 6.5 nm sample is in good agreement with calculations. Also, the slightly triangular shape of the plateaus in figure 5 implies a 3D emitter.

In these structures, accurate spectroscopy of the states can be performed. A caveat to the spectroscopy is that the nominal as-grown epitaxial thicknesses are not sufficient to perform this spectroscopy, as monolayer fluctuations between devices gives error in the epitaxial dimensions. Figure 6 illustrates the representative fluctuations observed for randomly selected devices from this set of samples. These fluctuations can be accounted for by quantum well and barrier thickness variations, and can serve as a calibration of the device dimensions when

compared to a Thomas–Fermi self-consistent band structure model using the resonant peak positions as a fitting parameter [17].

Using these determined epitaxial dimensions, the band structure model can be used to determine at what bias states of given energy line up with the Fermi level. For a donor bound state, the binding energy is dependent on the quantum well thickness, barrier height and the position of the donor in the well. For a centrally located donor (maximum binding energy), donors will have $\sim 15 \text{ meV}$ binding energy. This corresponds to an applied voltage (forward bias) of 0.15 V for the device of figure 5. At this bias, numerous overlapping conductance peaks are observed for most devices. However, there often appear distinct lower-lying (in voltage) peaks, such as shown in figure 5. The observed peak position for the lowest-voltage peak in figure 5 corresponds to a binding energy of $\sim 35 \text{ meV}$, significantly larger than the expected maximum binding energy. Similar fittings of numerous other epitaxial structures give excess binding energies of $\sim 10\text{--}20 \text{ meV}$ in most cases, although no excess energy has been observed for a few small-area devices. We suggest two possible mechanisms for this apparent increase in binding energy. The first is that some of the donors reside in regions of the quantum well where monolayer fluctuations have enlarged the well, thus depressing the local quantum well energy state. This is consistent with the observation of peak positions shifting to lower voltages as the device area is increased, as a larger statistical number of fluctuations are sampled. The second possibility is that some states are donor complexes with larger binding energy. However, this is not consistent with the observed peak position shifts of this spectrum.

We have also been able to characterize the temperature dependence of the linewidths of these states. Figure 7 shows the experimentally measured linewidth of various isolated peaks in the 6.5 nm sample as a function of

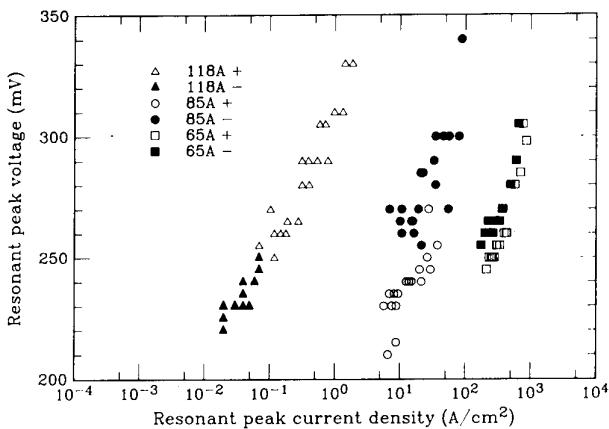


Figure 6. Resonant peak voltage position versus resonant peak current density for a random sampling of devices for the three barrier thicknesses, nominally 11.8 nm, 8.5 nm and 6.5 nm. Both positive (+) and negative (-) polarities are shown. $T = 77 \text{ K}$.

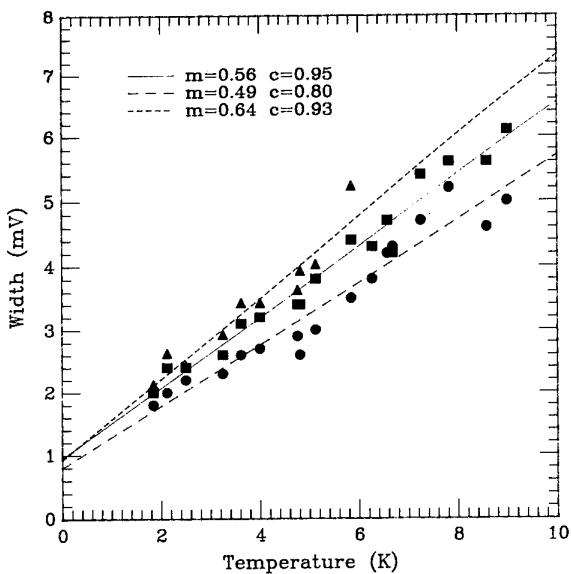


Figure 7. Measured linewidth (FWHM, in voltage) of various isolated peaks in the 6.5 nm sample as a function of temperature. The slope and $T = 0$ intercept, determined by least-squares fitting, are labelled for each distinct peak. The triangles, squares and circles correspond to the short-dash, full and broken lines, respectively.

temperature. The three lowest-lying peaks are shown, which translates to ‘excess’ binding energies ranging from ~ 20 to ~ 5 meV. (Higher-voltage peaks are not isolated, and thus are not presented.) The expected temperature broadening of the peaks simply corresponds to the broadening of the Fermi level, for which $\Delta V_{G,\text{FWHM}}(T) = 3.5 kT\alpha^{-1}$, where α is the voltage to energy conversion ratio in the well. Assuming $\alpha = 0.5$, a good approximation for a symmetric epitaxial structure under low bias, $\Delta V(T) = 0.6 \text{ mV K}^{-1}$. Figure 7 shows slopes of $\sim 0.6 \text{ mV K}^{-1}$ for the three peaks, as does a similar study for the 8.5 nm sample. The $T = 0$ intercept should give the intrinsic linewidth of the states, assuming that the external noise is very much less than the linewidth (a measurement of the linewidth under these conditions will be reported elsewhere).

Finally, we have observed for the first time that some plateaus exhibit ‘oscillations’ in current magnitude within a plateau. Figure 8 illustrates this effect, also presented as a function of temperature. This device is from the 8.5 nm sample; a comparison of figures 8 and 5 is representative of the observed variation of the oscillation effect between nominally identical devices. The broadening of the initial step rise is just the Fermi level broadening. The oscillations are relatively temperature insensitive with respect to the Fermi level broadening, indicating that they are not due to tunnelling through additional parallel impurity paths. The mechanism for the oscillations is not yet clear, although we suggest quantum interference with either structural defects, interface defects or other impurities as a possible explanation.

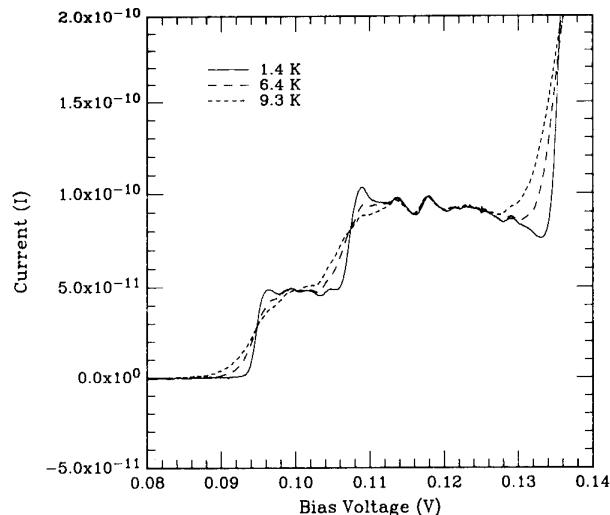


Figure 8. ‘Oscillation’ effect for a 8.5 nm barrier $8 \mu\text{m}^2$ device, as a function of temperature.

3. Summary

Transport in laterally confined heterostructures will generally exhibit a superposition of quantum dot, Coulomb blockade and impurity states. We have presented here the variety of effects observed, and techniques to distinguish between them. An effect which can mimic artificially imposed charge or energy quantization is impurity tunnelling, such as single-electron tunnelling through a *single* donor bound state which we have observed in simple non-confined heterostructures.

Acknowledgment

This work was supported by NSF DMR-9112497.

References

- [1] Reed M A, Randall J N, Aggarwal R J, Matyi R J, Moore T M and Wetsel A E 1988 *Phys. Rev. Lett.* **60** 535
- [2] Reed M A, Randall J N, Luscombe J H, Frenslay W R, Aggarwal R J, Matyi R J, Moore T M and Wetsel A E 1989 *Advances in Solid State Physics* vol 29 ed U Rössler (Braunschweig: Vieweg) p 267
- [3] Tarucha S, Hirayama Y, Saku T and Kimura T 1990 *Phys. Rev. B* **41** 5459
- [4] Tewordt M, Law V J, Kelly M J, Newbury R, Pepper M, Peacock D C, Frost J E F, Ritchie D A and Jones G A C 1990 *J. Phys.: Condens. Matter* **2** 8969
Tewordt M, Ritchie D A, Syme R T, Kelly M J, Law V J, Newbury R, Pepper M, Frost J E F, Jones G A C and Stobbs W M 1991 *Appl. Phys. Lett.* **59** 1966
- [5] Su B, Goldman V J, Santos M and Shayegan M 1991 *Appl. Phys. Lett.* **58** 747
- [6] Ramdane A, Faini G and Launois H 1991 *Z. Phys. B* **85** 389
- [7] Asahi H, Tewordt M, Syme R T, Kelly M J, Law V J, Frost J E F, Ritchie D A, Jones G A C and Pepper M 1991 *Surf. Sci.* **267** 388

[8] Guéret P, Blanc N, Germann R and Rothuizen H 1992 *Phys. Rev. Lett.* **68** 1896

[9] Su B, Goldman V J and Cunningham J E 1992 *Science* **255** 313; 1992 *Phys. Rev. B* **46** 7644

[10] Tewordt M, Martin-Moreno L, Nicholls J T, Pepper M, Kelly M J, Law V J, Ritchie D A, Frost J E F and Jones G A C 1992 *Phys. Rev. B* **45** 14407

[11] Blanc N, Guéret P, Germann R and Rothuizen H 1993 *Physica B* **189** 135

[12] Averin D V, Korotkov A N and Likharev K K 1991 *Phys. Rev. B* **44** 6199

[13] Dellow M W, Beton P H, Langerak C J G M, Foster T J, Main P C, Eaves L, Henini M, Beaumont S P and Wilkinson C D W 1992 *Phys. Rev. Lett.* **68** 1754

[14] Main P C, Beton P H, Dellow M W, Eaves L, Foster T J, Langerak C J G M, Henini M and Sakai J W 1993 *Physica B* **189** 125

[15] Sakai J W, Fromhold T M, Beton P H, Eaves L, Henini M, Main P C, Sheard F W and Hill G 1993 *Phys. Rev. B* **48** 5664

[16] Preliminary results have appeared in Deshpande M R, Dekker N H, Sleight J W, Hornbeck E S, Reed M A, Matyi R J, Kao Y C, Fernando C L and Frensley W R 1993 *Proc. IEEE/Cornell Conf. on Advanced Concepts in High Speed Semiconductor Devices and Circuits* (2–4 August 1993) (New York: IEEE)

A similar result was subsequently reported in Sakai J W, La Scala N Jr, Main P C, Beton P H, Foster T J, Geim A K, Eaves L, Henini M, Hill G and Pate M A 1993 *6th Int. Conf. on Modulated Semiconductor Structures* (23–27 August 1993)

[17] Reed M A, Frensley W R, Duncan W M, Matyi R J, Seabaugh A C and Tsai H-L 1989 *Appl. Phys. Lett.* **54** 1256

[18] Kopley T E, McEuen P L and Wheeler R G 1988 *Phys. Rev. Lett.* **61** 1654

Tunnelling spectroscopy of 0D states

J Smoliner[†], W Demmerle[‡], E Gornik[†], G Böhm[‡] and G Weimann[‡]

[†] Institut für Festkörperferelektronik, Floragasse 7, A-1040 Wien, Austria

[‡] Walter Schottky Institut, TU München, D-8046 Garching, Germany

Abstract. We have studied tunnelling processes between a multiple quantum dot (MQD) system and a two-dimensional electron gas (2DEG) system, which are realized on a GaAs-AlGaAs-GaAs heterostructure. Using a transfer Hamiltonian formalism it is shown that the tunnelling probability for transitions between a zero-dimensional (0D) and a two-dimensional (2D) state strongly depends on the quantum dot potential profile. In the case of a square well potential, only the resonance of the ground state is pronounced significantly, whereas for a cosine-shaped quantum dot potential profile a multitude of resonance structures is caused by each 0D state. From our experimental results we conclude that the potential of the quantum dots is best described by a cosine-shaped profile. In addition, the subband spacings and the extent of the wavefunctions of the individual subbands are also determined directly.

1. Introduction

Recent advances in microfabrication technologies such as lithography, etching and epitaxial regrowth have made it possible to reduce the dimensions of semiconductor devices and also the dimensionality of electronic states therein. Therefore it is possible to confine the motion of electrons in all three spatial directions in a way that their energy is completely quantized. The first experiments on zero-dimensional (0D) systems in semiconductors were mainly concentrated on the investigations of their optical properties using photoluminescence [1, 2] and cathode luminescence spectroscopy [3]. By measuring the differential capacity of a multiple quantum dot (MQD) system as a function of the gate voltage or an external magnetic field, the 0D density of states is directly reflected in the experimental results [4, 5]. At high magnetic fields, a splitting of the energy levels in surface states and bulk-like Landau levels was observed [6]. Most recently, capacitance measurements have been performed on single quantum dots [7]. Changes in the capacitance between a metal electrode and the discrete states of the quantum dot can be assigned to tunnelling processes of single electrons. Using far-infrared (FIR) transmission spectroscopy, intraband transitions between the 0D states of quantum dots, realized on InSb, were also observed directly [8]. Similar experiments were carried out on a GaAs-Al_xGa_{1-x}As system [9] and on silicon MOS structures [10]. On small structures with few electrons, non-local interactions and single-particle effects have also to be taken into account [11].

Tunnelling through 0D states in ultra small, pillar-shaped double-barrier resonant tunnelling diodes (DBRTDS) was observed for the first time by Reed *et al* [12]. Resonance structures in the tunnelling current were assigned to resonant tunnelling processes between 1D states in the emitter region and the 0D states inside the

quantum well defined by the two barriers. An anharmonic oscillator model was used to simulate the experimental results [13]. Further theoretical investigations of these tunnelling processes predict a strong coupling between the 1D subbands in the contact region and the 0D states inside the quantum dot in order to explain the fine structure of the current–voltage characteristic [14]. In similar experiments, based on asymmetric DBRTDS, charging effects were also observed [15, 16]. Besides etching, focused ion beam implantation can also be used to define structures containing quantum dots [17]. In most of the experiments cited above a harmonic-oscillator-shaped potential profile is used to describe the results in a sufficiently exact way. Using a shallow etching process in combination with a side gate, it is possible to adjust the 0D confinement by varying the gate voltage. The observed resonance structures in the tunnelling current through these devices were assigned to Coulomb blockade effects [18] as well as to ionized donor atoms [19]. Based on similar considerations, using triple-barrier RTDS as a starting point, tunnelling processes in coupled quantum dot structures have been investigated both experimentally [20] and theoretically [21].

In this paper, we investigate tunnelling processes between a two-dimensional electron gas (2DEG) system and a MQD system. The dot structures are regularly arranged on a square grid on the sample surface, the plane of which is parallel to the 2DEG system beneath. As both systems are separated by a 200 Å thick Al_xGa_{1-x}As barrier, electrons can tunnel vertically between 2D subbands and the 0D states of the quantum dots. It is important to note that in the 2DEG system the energy of motion in the direction of the tunnelling current is quantized. This geometrical circumstance results in selection rules which depend strongly on the profile of the quantum dot potential. As reported previously, in connection with 1D–2D tunnelling processes [22], we

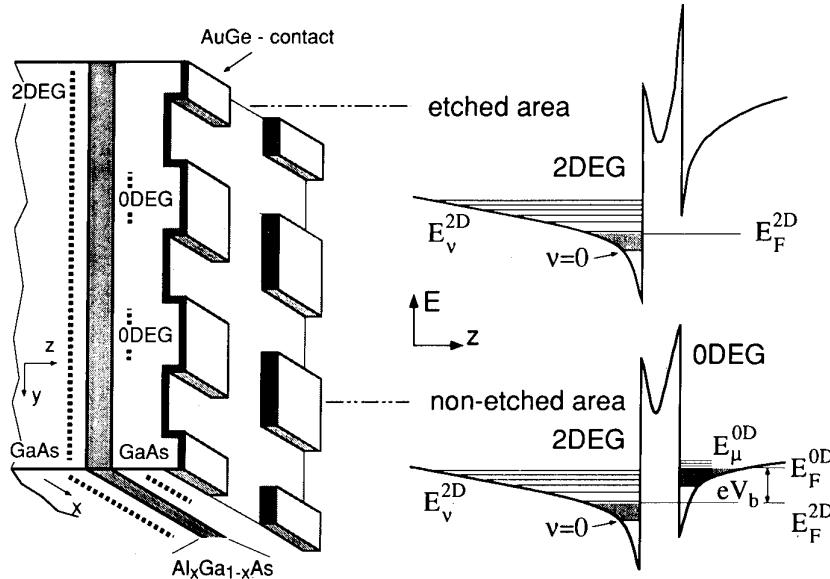


Figure 1. A schematic view of the sample is shown on the left-hand side. The corresponding conduction band profile for the etched and non-etched areas of the samples is shown on the right.

present a model based on the transfer Hamiltonian formalism in order to demonstrate this effect, and the influence of the quantum dot potential profile is discussed.

2. Sample preparation and processing

We have studied heterostructures consisting of a nominally undoped GaAs layer ($N_A < 1 \times 10^{15} \text{ cm}^{-3}$) grown on a semi-insulating substrate, followed by an Al_xGa_{1-x}As barrier of a total thickness of 200 Å (50 Å spacer, 50 Å doped ($N_D = 3 \times 10^{18} \text{ cm}^{-3}$), 100 Å spacer; $x = 0.36$), and an *n*-doped GaAs layer ($d = 800 \text{ Å}$, $N_D = 1.2 \times 10^{15} \text{ cm}^{-3}$). An additional GaAs cap layer was highly *n*-doped ($d = 150 \text{ Å}$, $N_D = 6.4 \times 10^{18} \text{ cm}^{-3}$). This structure provides an inversion layer at the lower GaAs/Al_xGa_{1-x}As interface, containing several 2D subbands, and an accumulation layer at the upper Al_xGa_{1-x}As/GaAs interface, separated by a barrier of only 200 Å. From Shubnikov-de Haas measurements it was deduced that in both 2DEG systems only one subband is occupied, having electron concentrations of $n_s^{\text{inv}} = 6.0 \times 10^{11} \text{ cm}^{-2}$ and $n_s^{\text{acc}} = 5.5 \times 10^{11} \text{ cm}^{-2}$ respectively.

For sample processing, bar-shaped mesas were etched and ohmic AuGe contacts were aligned to both channels. Then, a holographic photoresist grid with a period of $a = 350 \text{ nm}$ was fabricated on the mesas, using a UV laser interference pattern and a double exposure technique. In order to deplete the accumulation layer in the uncovered regions, the dots were etched wet chemically around 300 Å deep into the GaAs cap layers. The remaining islands have a diameter of $\approx a/2$, resulting in a three-dimensional confinement of the carriers in the accumulation layer. Finally, AuGe was evaporated on the total area of the tunnelling contact, establishing ohmic

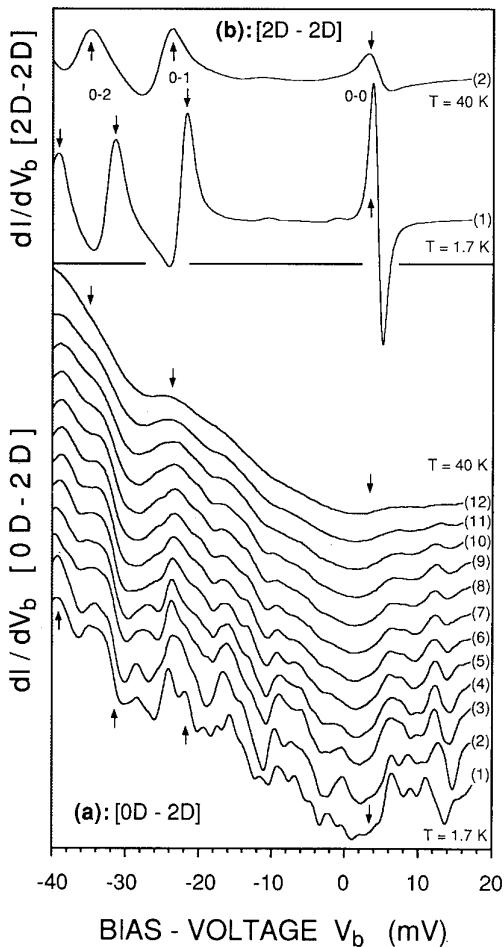
contacts to the quantum dot system. The etched areas between the dots remain surface depleted. The sample geometry is shown schematically in figure 1(a). From Shubnikov-de Haas measurements we are sure that the etch profile is not deep enough to induce a significant electron density modulation in the lower channel. On the other hand, a leakage tunnelling current can still flow directly from the AuGe contacts on the etched areas into the 2D channel. However, this non-resonant contribution to the tunnelling current is considerably smaller than the 0D to 2D resonant tunnelling current. Finally, the GaAs layers around the top tunnelling contact were removed selectively, yielding independent contacts to both the MQD system and the 2DEG system. As a result, we have a system where quantum wires and a two-dimensional electron gas are independently contacted, although separated only by a barrier of 200 Å.

By applying a voltage V_b , the 0D states are shifted energetically by $\Delta E = e\Delta V_b$ with respect to the 2DEG, since this external electric field drops completely across the potential barrier [23]. The electronic states on both sides of the barrier are therefore directly reflected in the tunnelling current. A negative bias voltage $V_b < 0$ corresponds to tunnelling processes from a 0D state of the MQD system into a 2D subband of the inversion channel. The band structure of the sample, which illustrates this situation, is shown in figure 1 for both the etched and the non-etched regions (upper and lower part respectively). All measurements of the current-voltage characteristic $I(V_b)$ and its first derivative, dI/dV_b , were made using a four-terminal conductance bridge [24] in order to compensate series resistances inside the contacts and the inversion channel. A modulation frequency of 22 Hz and modulation voltage of 0.1 mV were chosen to achieve a high resolution and to avoid parasitic capacitance perturbations.

3. Experimental results

The experimental results are plotted in figure 2. The lower part (a) of this figure shows the dI/dV_b characteristics of the nanostructured (0D–2D) sample, where the temperature is varied between $T = 1.7$ K (curve (1)) and $T = 40$ K (curve (12)). For reference purposes, the dI/dV_b characteristics of a non-nanostructured sample are plotted in the upper part (b) of figure 2 for the two temperature values $T = 1.7$ K (curve (1)) and $T = 40$ K (curve (2)). The comparison of the two characteristics (a(1)) and (b(1)) shows that the nanofabrication process leads to a multitude of new resonance structures which exist within the whole voltage range considered. Especially at small negative bias voltages new resonance structures can be observed, in a regime where the characteristic (b(1)) is absolutely flat. Concerning the amplitudes and the positions of the new peaks there exists no obvious correlation between the 2D–2D and the 0D–2D tunnelling characteristic. Nevertheless it has to be stressed that all resonance structures are fully reproducible within the investigated bias voltage regime.

All resonance peaks of the 0D–2D dI/dV_b characteristic, however, show a strong dependence on the temperature. Above $T = 4.2$ K (curve a(2)), only about half of the resonances can still be resolved. A further increase of the temperature results in a monotonic broadening of all resonance structures, accompanied by a monotonic decrease of peak amplitudes. For bias voltages $V_b < 0$ one can observe a common shift of all structures to more negative values of V_b , whereas the peak positions at positive bias voltage remain almost unchanged at temperatures $T > 4.2$ K. The subband resonances of the 2D–2D dI/dV_b characteristic (b(1)) behave in the same way. With increasing temperature the peak amplitudes decrease rapidly, whereas the linewidth increases simultaneously. The shift of these resonance structures is due to the thermally activated occupation of the first excited subband in the inversion channel which causes a self-consistent modification of the potential profile. In both parts (a) and (b) of figure 2 the positions of the 2D–2D subband resonances are marked by arrows. Comparison of the two high-temperature characteristics (a(12)) and (b(2)) shows that the positions of the still resolvable resonances correspond to each other.



4. 0D–2D tunnelling selection rules

In analogy to the determination of the selection rules for 1D–2D tunnelling processes which were recently published [22], we calculate in this section the tunnelling probability for transitions between the subbands of the 2DEG system and the fully quantized states of the MQD system. The model is based on the transfer Hamiltonian formalism, a method which is ideally suited for taking states of different dimensionality in the electrodes of the tunnel diode into account. In order to describe the states in the MQD system it is sufficient to consider one isolated quantum dot. Because of the distance between two adjacent quantum dots and the height of the potential modulation (approximately 50 meV) an interaction between two degenerate 0D states can be definitely neglected. In addition, the ‘motion’ in the direction of the tunnelling current (z) is expected to be uncoupled from the remaining coordinates. For this reason, a cylindrical coordinate system is ideally suited to describe the electronic properties of the 0D system investigated here. The wavefunctions are therefore characterized by a function $\Psi_A^{0D}(\rho, \varphi, z; n_A, m_A)$, where n_A represents the radial and m_A the azimuthal quantum number. As the unstructured accumulations layer (index A) exhibits only one subband, these two quantum numbers are sufficient to describe the 0D states in the MQD system. The wavefunctions of the subbands in the inversion layer (index I) are given by $\Psi_{I,v}^{2D}(\rho, \varphi, z; m_I)$. Here, v stands for the 2D subband index; m_I takes the degeneration of E_{\parallel} (energy of the motion parallel to the interface) into account. As a coupling between the coordinates is neglected, it is possible to use the following separation

Figure 2. (a) Measured dI/dV_b curves in the temperature range between 1.7 K and 40 K. (b) For comparison, dI/dV_b curves for unstructured samples are also shown.

ansatz

$$\begin{aligned} \text{2DEG: } \Psi_{I,v}^{2D} &= \frac{1}{\sqrt{A_{\rho\varphi}}} \psi_{I,v}(z) \frac{1}{\sqrt{2\pi}} \exp(im_I\varphi) J_{m_I}(k_{||}\rho) \\ \text{0DEG: } \Psi_A^{0D} &= \frac{1}{\sqrt{A_{\rho\varphi}}} \psi_A(z) \frac{1}{\sqrt{2\pi}} \exp(im_A\varphi) \Phi_{n_A, m_A}(\rho). \end{aligned} \quad (1)$$

The area $A_{\rho\varphi}$ represents the normalization constant in both systems. The total energy is therefore given by

$$\begin{aligned} \text{2DEG: } E^{2D}(v_I, m_I; k_{||}) &= E_{v_I} + \frac{\hbar^2 k_{||}^2}{2m^*} \\ \text{0DEG: } E^{0D}(n_A, m_A) &= E_{n_A, m_A}. \end{aligned} \quad (2)$$

In the inversion channel, E_{v_I} stands for the energy of the subband edges, whereas $k_{||}$ is the wavevector of the motion parallel to the interface. The total energy in the 2DEG system is degenerate with respect to the quantum number m_I^2 . The energy eigenvalues of the quantum dot are characterized by the quantum numbers n_A and m_A ; the energy of the z ground state (non-nanostructured accumulation channel) is already included.

If a voltage V_b is applied to the sample and drops completely across the tunnelling barrier, the energy levels of both systems are shifted energetically by $\Delta E = e\Delta V_b$ with respect to each other. The conservation of the total energy during the tunnelling process leads to the first selection rule

$$E^{2D}(v_I, m_I; k_{||}) = E^{0D}(n_A, m_A) - eV_b. \quad (3)$$

This condition has to be considered when calculating the transition probability for a tunnelling process between a 0D state of the quantum dot and the 2D subband of the inversion channel. The corresponding rate is given by

$$P_{AI} = \frac{2\pi}{\hbar} |M_{AI}|^2 \delta(E^{2D} - E^{0D} + eV_b). \quad (4)$$

In this equation, the δ -function represents the conservation of the total energy. Within the transfer Hamiltonian formalism [25], the matrix element $|M_{AI}|^2$ which is governed by the overlap of the single wavefunctions can be obtained by the following expression

$$\begin{aligned} M_{AI} &= \frac{-\hbar^2}{2m^*} \iint_S dx dy \left(\Psi_I^* \frac{\partial \Psi_A}{\partial z} - \Psi_A \frac{\partial \Psi_I^*}{\partial z} \right) \\ &= \frac{-\hbar^2}{2m^*} \underbrace{\left[\psi_{I,v}^* \frac{\partial \psi_A}{\partial z} - \psi_A \frac{\partial \psi_{I,v}^*}{\partial z} \right]_{z=z_b}}_{t_B} \\ &\quad \times \delta_{m_I, m_A} \times \underbrace{\int d\rho \rho J_{m_I}(k_{||}\rho) \Phi_{n_A, m_A}(\rho)}_{\langle J_m(k_{||}\rho) \rangle \langle \Phi_{n_A, m}(\rho) \rangle}. \end{aligned} \quad (5)$$

The first term in this equation, t_B , represents the overlap of the wavefunctions in the direction of the tunnelling current and corresponds to the transmission coefficient of

a single-barrier heterostructure. The δ -function in the middle of equation (5) guarantees the conservation of the angular momentum (quantum number m) during the tunnelling process. For the matrix element non-zero values can only be obtained for $m_I = m_A \equiv m$.

The overlap of the radial part of the wavefunctions is described by the last term in equation (5). The value of the corresponding matrix element is a function of the radial quantum number n_A , the (common) azimuthal quantum number m , and the wavevector $k_{||}$ which depends on the applied bias voltage. Its value, however, is totally fixed by demanding the conservation of the total energy (equation (3))

$$k_{||}(V_b) = \sqrt{\frac{2m^*}{\hbar^2} (E_{n_A, m} - E_{v_I} - eV_b)}. \quad (6)$$

The value of the tunnelling current is, beside the tunnelling probability, also determined by the number of 0D states which are located energetically between the two Fermi levels in the emitter and in the collector respectively. The total resonant tunnelling current is therefore given by

$$\begin{aligned} j_{\text{tot}} &= \frac{4\pi e}{\hbar} \sum_{n_A, m} \sum_{v_I} |M_{AI}|^2 \times [f_I(E^{2D}) - f_A(E^{0D} - eV_b)] \\ &= \frac{4\pi e}{\hbar} \sum_{n_A, m} \sum_{v_I} |t_B|^2 g_{n_A, m} |\langle J_m(k_{||}\rho) \rangle \langle \Phi_{n_A, m}(\rho) \rangle|^2 \\ &\quad \times [f_I(E^{2D}) - f_A(E^{0D} - eV_b)]. \end{aligned} \quad (7)$$

$|t_B|^2$ represents the single-barrier transmission coefficient, which is weakly exponential in the small voltage range used for our experiments and contains no structure as a function of V_b . The factor $g_{n_A, m}$ takes the possible degeneration of the 0D energy levels into account. The overlap matrix element correlates with the tunnelling probability and is in principle responsible for all resonance structures in the tunnelling current. The last term in equation (7) guarantees that tunnelling is possible only if the initial state is occupied and the final state is free (f_I and f_A are the Fermi distribution functions for the inversion and the accumulation channel respectively). Thus, all structure in the resonant tunnelling current will be caused by the behaviour of the overlap matrix element $I_{n_A, m}$ as a function of bias voltage.

Our calculations have shown that the value of the overlap matrix element $I_{n_A, m}$ not only depends on the quantum numbers n_A and m and on the bias voltage V_b , but also on the particular potential profile. In particular the boundary conditions of the confining potential play an important role. Self-consistent calculations of the potential profile for deep mesa etched, cylindrical-shaped structures [12] show that the central part of the quantum dot potential becomes flat with increasing doping concentration of the material, whereas the boundary of the confining potential is then characterized by steep sidewalls [26]. This is mainly due to the smaller depletion width at the open etched surfaces. The electronic properties of such systems are therefore reasonably well described by the means of a square well potential profile.

Shallow etched quantum dots, however, as they exist on our samples, are better described by parabolic or cosine-shaped potentials. In our system the electron concentration is rather high, and thus many 0D states are occupied and contribute to the tunnelling current. Deviations from a parabolic potential are therefore expected for states higher in energy, which results in a lifting of the degeneration of some energy levels, leading to a more complex tunnelling spectrum. In order to take a realistic potential situation into account, we chose a cosine-shaped quantum dot potential profile with radial symmetry, which is analytically given by the following expression

$$V_{\text{dot}}(\rho) = V_{\text{mod}} \left[\frac{1}{2} + \frac{1}{2} \cos \left(\frac{\pi}{R_{\text{dot}}} (\rho - R_{\text{dot}}) \right) \right]$$

with $\rho \leq R_{\text{dot}}$. (8)

In order to analyse the influence of the shape of the confining potential on the probability for tunnelling processes between 0D and 2D states in the present structure, we first discuss the results for a radially symmetric square well potential of finite height. The parameters for this model potential are chosen to obtain comparable quantization energies with respect to the cosine-shaped potential discussed above.

In figure 3 we have plotted the wavefunction overlap as a function of bias voltage for the lowest four quantum numbers of $n_A = 1-4$. The overlap integral $I_{n_A, m}$ of the $n_A = 0$ ground state shows only a sharp resonance structure at $V_b = V_0$ for $m = 0$ (upper part of figure 3). In the regime $V_b > V_0$ no resonant tunnelling process is allowed, since the total energy cannot be conserved during the transition. For $V_b < V_0$, the value of $I_{0,0}$ drops exponentially towards zero within a certain voltage range. The functions $I_{0,m}(V_b)$ corresponding to the states with higher angular momentum quantum number ($m \geq 1$) also show a maximum in this regime; however, the linewidth of the functions increases and their intensity decreases drastically as m becomes larger. The overlap integrals of the radial quantum numbers $n_A = 1, 2$ and 3 are also plotted in figure 3 for different values of m . In all cases, a sharp resonance structure can only be obtained for the particular $m = 0$ states. The structures which are due to higher values of m always occur at more positive bias voltages than for $m = 0$. In addition, all overlap integrals for larger n values ($m = 0$) have a large and relatively broad peak at the voltage position V_0 .

To understand this behaviour, one has to start from equation (6). According to this equation, the wavevector $k_{||}$ of the tunnelling electrons is tuned by the applied bias voltage as all other parameters in this equation are constant. Analysing the peak positions ($m = 0$) quantitatively, one can see that each peak in the overlap integral is due to a situation where the expression $i\pi/k = w$ is equal to the width of the dot, w (i is an integer). In other words, peaks in the wavefunction overlap are always expected, when an integer multiple of half of the wavelength of the tunnelling electron, $\lambda/2$, fits into the dot, where $k = 2\pi/\lambda$. Note that this explains why

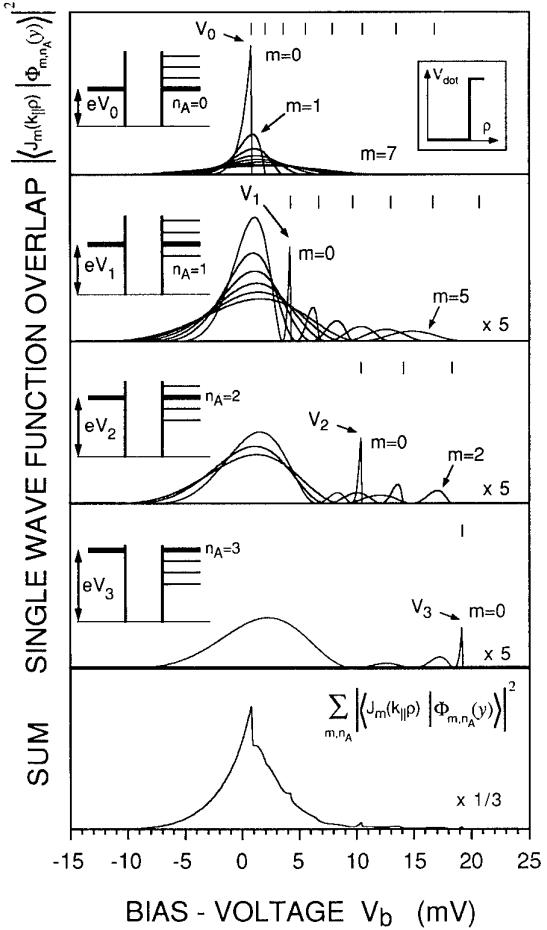


Figure 3. Calculated wavefunction overlap between the 2D state and the lowest three 0D states for a quantum dot with rectangular potential of finite height. The sum over all 0D subbands is also shown.

the resonance peak for the overlap between the lowest 0D and the 2D subband is not exactly at the position where the subbands are aligned, but at somewhat more negative bias voltages. Here, the voltage where the width of the dot is equal to $w = \pi/k$ simply corresponds to the zero-point energy in the dot.

In general, the transition probability of each (radial) 0D state exhibits many different sharp and also broad resonance structures which leads to a very complex total tunnelling probability, as one has to sum over all states in the relevant energy range. The lowest part of figure 3 shows the sum of all single-wavefunction overlap integrals $I_{n_A, m}(V_b)$ for the rectangular potential. The resonance structures caused by the $m = 0$ states can still be clearly resolved, but no other maxima of $I_{n_A, m}(V_b)$ can be definitely assigned to a particular tunnelling process since they are too weak and contributions of too many 0D states interfere. However, all functions $I_{n_A, m}(V_b)$ have a finite value in the vicinity of $V_b \approx V_0$ where they contribute to the total tunnelling probability. This effect leads to a global maximum of $\sum I_{n_A, m}$ around the position of the 0D ground state resonance, which will become

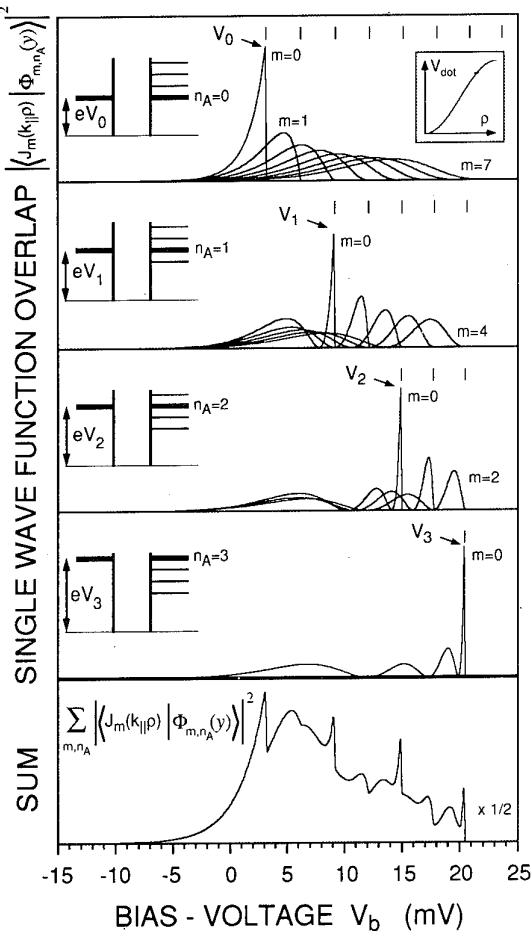


Figure 4. Calculated wavefunction overlap between the 2D state and the lowest three 0D states for a quantum dot with cosine-shaped potential of finite height. The sum over all 0D subbands is also shown.

important when discussing the temperature behaviour of our experimental results.

We now compare the above findings with the results for a cosine-shaped potential, which are shown in figure 4. At first sight the shape of the overlap integrals is the same as for the rectangular potential, but somewhat broadened. If we look at the peak positions for the overlap integrals of the higher subbands we can see two differences. First, the peaks are shifted with increasing subband index n and second, the peak structure is much more pronounced, especially if we look at the sum of the overlap integrals at the bottom of figure 4. The shift of the peaks is understood by the following arguments: as mentioned above, peaks in the wavefunction overlap are always expected each time $i\pi/k = w$. In a cosine-shaped potential the extension of the subbands increases with increasing subband index. This explains the shift of the peak positions to different bias voltages, which correspond to different k values. The more pronounced peak structure is due to the different shape of the wavefunctions in the cosine-shaped potential, since the overlap integral can be regarded as a Fourier transform of the quantum dot states. In a finite rectangular potential the wave-

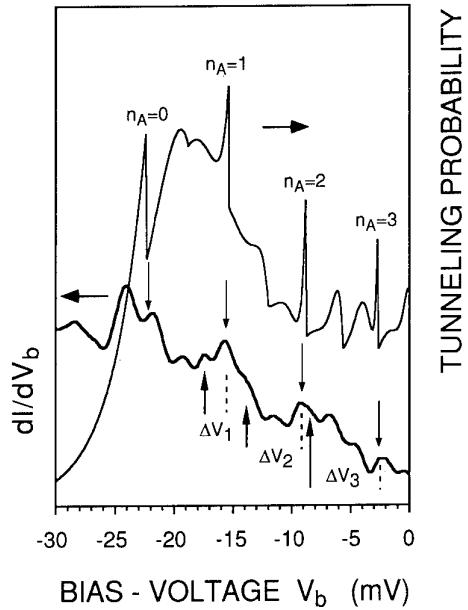


Figure 5. Comparison between the calculated tunnelling probability and the measured dI/dV_b curve at 1.7 K. The downward arrows mark the $n = 0-3$ resonance peaks. The upward arrows mark the $i = 1$ peak positions for each subband.

functions are sinusoidal, which means that the Fourier transform of the wavefunction looks like a δ -function. For the cosine-shaped potential, the wavefunctions are roughly harmonic-oscillator-like, which means that they contain ‘higher frequencies’ and the corresponding Fourier transform has a rich structure in it.

5. Comparison with experiment

Figure 5 shows a comparison between the calculated tunnelling probability and the dI/dV_b curve measured at 1.7 K. As one can see, the experimental results agree well with the calculated peak positions obtained for a cosine-shaped model potential with $R_{dot} = 62.5$ nm and $V_{mod} = 38$ meV. Not only the peak positions but also the larger modulation of the measured dI/dV_b curve are explained by the model. These larger maxima, on which the small peaks are superimposed, are due to contributions of many 0D states, which is consistent with the observed temperature behaviour of the peak amplitudes in dI/dV_b . The small peaks smear out very quickly with increasing temperature, but the larger maxima stay clearly present up to $T \approx 15$ K since they are due to the interference of many 0D states. The weakly exponential background is not reproduced by the model calculation, since the transmission coefficient of the barrier was assumed to be constant for simplicity. Note that, strictly speaking, the calculated tunnelling probability has to be compared with the current–voltage characteristics. In $I(V_b)$, however, the structures are too weak to be resolved, and thus the comparison with dI/dV_b is justified.

We now assign the sharp structure in the total

tunnelling probability which are due to the sharp peaks in the wavefunction overlap integrals for $n_A = 0, 1, 2, 3$, with single peaks in the tunnelling characteristics (marked in figure 5 with downward arrows) and determine the subband spacings between the lowest three 0D subbands in the quantum dot. If we take into account that to a good approximation the relative energy shift of the 2D and 0D states is equal to $e\Delta V_b$, the subband spacings of the lowest three subbands are determined as $\Delta E_{01} \approx 7$ meV, $\Delta E_{12} \approx 6$ meV, $\Delta E_{23} \approx 5$ meV. These relatively high subband spacings are consistent with the parameters used for sample preparation, as the dots were etched rather deep into the GaAs on this sample. Note that the decreasing subband spacings with increasing subband index clearly show that the potential in the dot is not rectangular or parabolic and that a cosine-shaped potential is the most realistic description of the situation.

After we have assigned the $n_A = 0, 1, 2, 3$, resonances in the tunnelling characteristics, which make it possible to determine the 0D subband spacing, we also identify the other peaks in dI/dV_b and estimate the extent of the wavefunction for each 0D subband. For this purpose we use our finding that the most pronounced structures in the calculated tunnelling probability are due to situations where a multiple integer of half of the electron wavelength matches the extent of the 0D state in the corresponding subband. All other structures, for example the peaks generated by the higher quantum number of m , are much smaller, interfere with each other and form the large modulation of the tunnelling probability, which persists to higher temperatures.

We now apply the simple relation $w = i\pi/k$ to estimate the extent of the wavefunctions in the $n = 1-3$ subbands. It must be pointed out that this relation between the width of the potential, w , and the wavevector k , only yields precise results for rectangular potentials. For cosine-shaped potentials, it can at least be used to check the order of magnitude of w for the lowest 0D subbands.

To determine the value of k , we first have to find the position where $k = 0$ for a tunnelling process into the n th 0D subband. For tunnelling processes into the $n_A = 3$ subband, peaks in the tunnelling characteristics occur for values of i in the range between $i = 1$ and $n + 1$. i is equal to 4 at the required position, which is marked by $n_A = 3$. If we ignore the zero-point energy of the 0D system for simplicity, the voltage position for $k = 0$ is approximately at the $i = 1$ peak position, which is marked by the upward arrows for all 0D subbands. Thus, the voltage spacing ΔV_3 between the $n_A = 3$ peak and the corresponding $i = 1$ peak can be used to determine k from equation (6), which then yields a value of $w = 120$ nm of the $n_A = 3$ subband. For $n_A = 2$ and $n_A = 3$ this procedure yields values of 102 nm and 94 nm of the $n_A = 2$ and $n_A = 1$ subbands respectively, which is in good agreement with the geometrical diameter of the dot of $d_{geo} \approx 160$ nm if the surface depletion is taken into account. Note that the extent of the $n_A = 0$ subband cannot be determined in this simple way, since the $k = 0$ position for the lowest 0D subband is not reflected directly in the tunnelling

characteristics. This also means that the zero point energy in the 0D system cannot be determined in the present experiment.

The good agreement between theory and experiment shows that tuning the wavevector of the tunnelling electron by the applied bias voltage really enables us to determine the extent of the wavefunction in the 0D states directly from the peak positions in the tunnelling characteristics. Thus, this sampling of the wavefunction by tuning the k vector of the tunnelling electron can be regarded as wavefunction (Fourier transform) spectroscopy.

6. Summary

In summary, we have studied tunnelling processes between a multiple quantum dot system and a two-dimensional electron gas system. The tunnelling probability for transitions between a 0D and a 2D state strongly depends on the quantum dot potential profile. Square well potentials only lead to few and narrow resonances, whereas for a cosine-shaped quantum dot potential profile a multitude of resonance structures is caused by each 0D state. This model turned out to be the most realistic description of the situation in our samples. As the main structures in the tunnelling probability always occur when a multiple integer of half of the wavelength of the tunnelling electron matches the extent of the corresponding 0D state, both the subband spacing and the extent of the 0D states were determined directly from the tunnelling characteristics. In the present experiments, the sampling of the wavefunction by changing the k vector of the tunnelling electron through the applied bias voltage can be regarded as *wavefunction spectroscopy*.

Acknowledgment

This work was partially sponsored by Oestereichische Nationalbank project no. 4874.

References

- [1] Reed M A, Bate R T, Bradshaw K, Duncan W M, Frensel W R, Lee J W and Shih H D 1986 *J. Vac. Sci. Technol. A* **4** 358
- [2] Temkin H, Dolan G J, Panish M B and Chu S N G 1987 *Appl. Phys. Lett.* **50** 413
- [3] Cibert J, Petroff P M, Dolan G J, Pearton S J, Gossard A C and English J H 1986 *Appl. Phys. Lett.* **49** 1275
- [4] Smith T P III, Lee K Y, Knoedler C M, Hong J M and Kern D P 1987 *Phys. Rev. Lett.* **59** 2502
- [5] Lee K Y, Smith T P III, Arnot H, Knoedler C M, Hong J M, Kern D P and Laius S E 1988 *J. Vac. Sci. Technol. B* **6** 1856
- [6] Hansen W, Smith T P III, Lee K Y, Brum J A, Knoedler C M, Hong J M and Kern D P 1989 *Phys. Rev. Lett.* **62** 2168
- [7] Ashoori R C, Störmer H L, Weiner J S, Pfeiffer L N,

Pearton S J, Baldwin K W and West K W 1992 *Phys. Rev. Lett.* **68** 3088

[8] Sikorski Ch and Merkt U 1989 *Phys. Rev. Lett.* **62** 2164

[9] Liu C T, Nakamura K, Tsui D C, Ismail K, Antoniadis D A and Smith H I 1989 *Appl. Phys. Lett.* **55** 168

[10] Alsmeyer J, Batke E and Kothaus J P 1990 *Surf. Sci.* **229** 287

[11] Demel T, Heitmann D, Grambow P and Ploog K 1990 *Phys. Rev. Lett.* **64** 788

[12] Reed M A, Randall J N, Aggarwal R J, Matyi R J, Moore T M and Wetzel A E 1988 *Phys. Rev. Lett.* **60** 535

[13] Luban M and Luscombe J H 1990 *Appl. Phys. Lett.* **57** 61

[14] Bryant G W 1991 *Phys. Rev. B* **44** 12838

[15] Su B, Goldman V J and Cunningham J E 1992 *Science* **255** 313

[16] Tewordt M, Martin-Moreno L, Nichols J T, Pepper M, Kelly M J, Law V J, Ritchie D A, Frost J E F and Jones G A C 1992 *Phys. Rev. B* **45** 14407

[17] Tarucha S, Tokuda J and Hirayama Y 1991 *Phys. Rev. B* **43** 9373

[18] Guéret P, Blanc N, Germann R and Rothuizen H 1992 *Phys. Rev. Lett.* **68** 1896

[19] Dellow M W, Beton P H, Langerak C J G M, Foster T J, Main P C, Eaves L, Henini M, Beaumont S P and Wilkinson C W D 1992 *Phys. Rev. Lett.* **68** 1754

[20] Tewordt M, Asahi H, Law V J, Syme R T, Kelly M J, Ritchie D A, Churchill A, Frost J E F, Hughes R H and Jones G A C 1987 *Appl. Phys. Lett.* **50** 413

[21] Fong C Y, Nelson J S, Hemstreet L A, Gallup R F, Chang L L and Esaki L 1992 *Phys. Rev. B* **46** 9538

[22] Demmerle W, Smoliner J, Gornik E, Böhm G and Weimann G 1993 *Phys. Rev. B* **47** 13574

[23] Demmerle W, Smoliner J, Berthold G, Gornik E, Weimann G and Schlapp W 1991 *Phys. Rev. B* **44** 3090

[24] Christianell R and Smoliner J 1988 *Rev. Sci. Instrum.* **59** 1290

[25] Harrison W A 1961 *Phys. Rev.* **123** 85

[26] Luscombe J H, Bouchard A M and Luban M 1992 *Phys. Rev. B* **46** 10262

Lasing in lower-dimensional structures formed by cleaved edge overgrowth

Werner Wegscheider, Loren Pfeiffer, Marc Dignam, Aron Pinczuk,
Kenneth West and Robert Hull

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

Abstract. We have used the molecular beam growth technique, which we call cleaved edge overgrowth, to fabricate highly efficient lasers that operate in the 1D quantum limit. The active region of our laser consists of atomically precise quantum wires that form at the T-shaped intersections of 7 nm wide GaAs quantum wells grown along the [001] and, after an *in situ* cleave, along the [110] crystal axis. The origin of the quantum mechanical bound state is the relaxation of quantum well confinement at this intersection, which leads to an expansion of the electron and hole wavefunction into the larger available volume at the T-junction. The high degree of structural perfection achievable in this way allows the observation of stimulated optical emission from the lowest exciton state in optically pumped devices. The interpretation that the observed quantum wire response is due to exciton recombination is based on the near spectral constancy of the emission over almost three orders of magnitude in excitation power from low-power luminescence to a single-mode lasing line. The implied absence of bandgap renormalization effects suggests that the Mott density is never reached and indicates interesting new behaviour of excitons in 1D. In contrast, the quantum well photoluminescence peak indeed shifts to lower energies consistent with the notion that the 2D excitons ionize and a free electron–hole plasma forms.

1. Introduction

Fabrication and operation of a semiconductor quantum wire (QWR) laser has been a challenge to the photonic field for more than a decade. Just as the quantum well (QW) lasers (2D) have replaced conventional double-heterojunction devices (3D) due to their superior performance, further improvements in threshold current and modulation bandwidths as well as reduced sensitivity of threshold current and emission wavelength to ambient temperature are expected from QWR lasers (1D) [1–3]. The sharp peak in the density of states (DOS) at the band edge of 1D systems, as opposed to the square root and step-like DOS profile in 3D and 2D respectively, should lead to a variety of interesting optical properties such as enhanced optical nonlinearities [4], narrower optical gain spectra and higher differential gain [1]. In addition, increased exciton binding [5, 6], anomalously strong concentration of the oscillator strength on the lowest-energy exciton state [7] and exciton condensation [8] are predicted as a result of Coulomb interaction in the electron–hole system confined to 1D.

While several growth techniques are well suited to the production of planar multilayer systems, the fabrication of structures in which charge carriers are quantum confined in more than one dimension is a very challenging task. Methods suggested and attempted include lithographic definition combined with etching and regrowth [9], growth on non-planar [10] and vicinal substrates [11, 12], selective area deposition [13] and local interdiffusion

[14, 15]. The first lasers exhibiting signature of carrier confinement to one dimension in the optical emission spectra have been prepared by organometallic chemical vapour deposition (OMCVD) on V-grooved substrates [10]. However, the relative large size of the crescent-shaped QWRs fabricated by this technique (80–100 by 10 nm) results in the occupation of many 1D subbands, and therefore in optical properties differing little from those of 2D systems. In order to observe characteristic 1D effects and the peculiar features associated with 1D excitons predicted by theory [7, 8] the dimensions of the QWRs should be comparable to or smaller than the bulk exciton Bohr radius. Otherwise only the centre-of-mass motion of the excitons shows 1D character while the electron–hole relative motion remains unaffected by the confining potential. It is, therefore, necessary to employ a fabrication technique which produces QWRs with uniform and precisely controlled dimensions of less than about 10 by 10 nm in the GaAs/AlGaAs system.

The QWRs reported in this paper which led to the first demonstration of stimulated optical emission using 1D exciton recombination [16] were realized by the cleaved edge overgrowth (CEO) method. This molecular beam epitaxy (MBE) technique is capable of producing intersecting QWs with atomic control of thicknesses in two orthogonal directions as described in detail elsewhere [17, 18]. The conceptually simple and straightforward method consists of two growth steps separated by an *in situ* cleave. After conventional growth of a QW structure on a [001]-oriented substrate, overgrowth takes place on

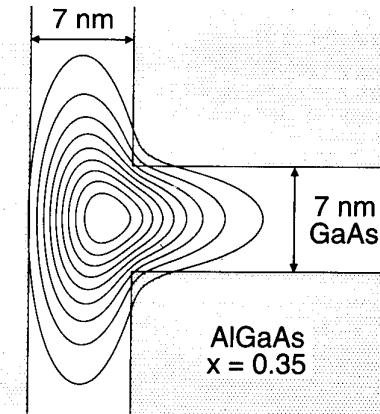


Figure 1. Contours of constant probability ($|\psi|^2 = 0.1, 0.2, \dots, 0.9$) for electrons confined at the T-intersection of two QWs.

the [110]-oriented sidewall which is exposed by the cleave. The post-cleave growth consists of another QW and a barrier layer. Figure 1 illustrates the physics of QWR formation at the T-shaped intersection of two QWs. The figure shows the cross section of such an intersection together with contours of constant probability for electrons. The origin of the quantum mechanical bound state is the relaxation of QW confinement at the intersection. While a classical particle would be unbound for the given T-shaped potential, the expansion of the wavefunction into the larger available volume at the junction results in a smaller kinetic contribution to the total energy of electrons and holes. Consequently, motion of the 1D carriers is limited to the line defined by the intersecting planes of the two QWs. In contrast to other QWR fabrication techniques which rely on introducing additional confinement of the carriers in a previously prepared 2D system, for example by partial removal of a QW layer and subsequent overgrowth with barrier material, the QWR states in our structure are energetically located below the ground state QW transitions. The existence of such confinement relaxation QWR states was experimentally demonstrated for the first time by Goñi *et al* [19].

2. Experiment

A schematic cross section of the QWR laser structure is shown in figure 2. The first MBE growth formed the layer structure to the right of the arrow marked ‘cleave’. It consists of a 1 μm $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ cladding layer followed by a 22-period $\text{GaAs}/\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$ multiple quantum well (MQW) structure with well and barrier thicknesses of 7 and 38 nm respectively, as illustrated in the magnified area, followed by a 3 μm $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ cladding layer. After growth of the layers to the left of the arrow marked ‘cleave’ 22 QWRs form at the T-intersections of the 7 nm wide QWs. The $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$ and $\text{Al}_{0.1}\text{Ga}_{0.9}\text{As}$ layers surrounding the QWRs serve as a T-shaped dielectric waveguide. Their purpose is to confine an optical mode

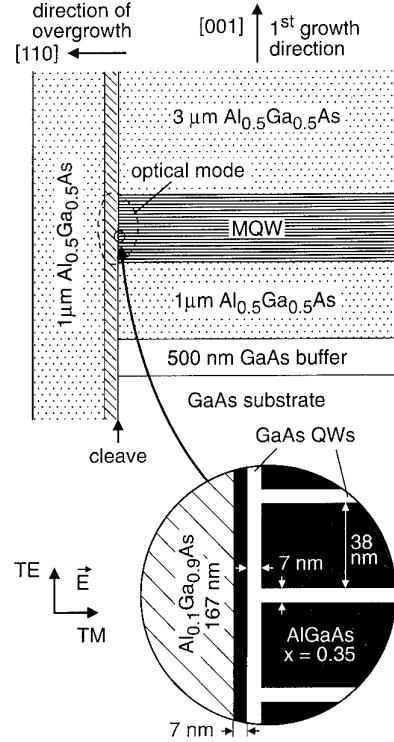


Figure 2. Cross section through the QWR laser structure. The broken line represents a contour plot of the optical mode at 10% of the maximum intensity.

to the vicinity of the QWR array (see the broken line in figure 2) as confirmed by waveguide calculations within the effective refractive index approximation. In this way a completely index-guided structure with an effective refractive index step of $\Delta n_{\text{eff}} = 0.029$ from the core of the T-shaped waveguide to the surrounding three-layer slab waveguides is obtained.

In order to achieve lasing in the 1D structures, mirrors were cleaved perpendicular to the axis of the QWRs, resulting in optical cavities 600 μm long. The cleaved mirrors were left uncoated for all experiments reported here. Low-temperature (1.7 K) continuous-wave excitation from either the (001) or the (110) surface of the samples was performed by focusing the output of a dye laser tuned to $\lambda = 775$ nm to a stripe of about 700 μm in length and 5 μm in width oriented parallel to the QWRs. At this wavelength significant light absorption occurs only in the GaAs QW and QWR layers. However, because the QWR volume is so small, light absorption takes place mainly in the QWs. Light transmitted through one of the mirrors was dispersed in a 0.85 m double monochromator and detected with a charge-coupled-device (CCD) camera.

3. Results and discussion

The high degree of structural perfection of the QWRs attainable by the CEO method is manifested by the planarity and abruptness of the interfaces along both growth directions as demonstrated in the transmission

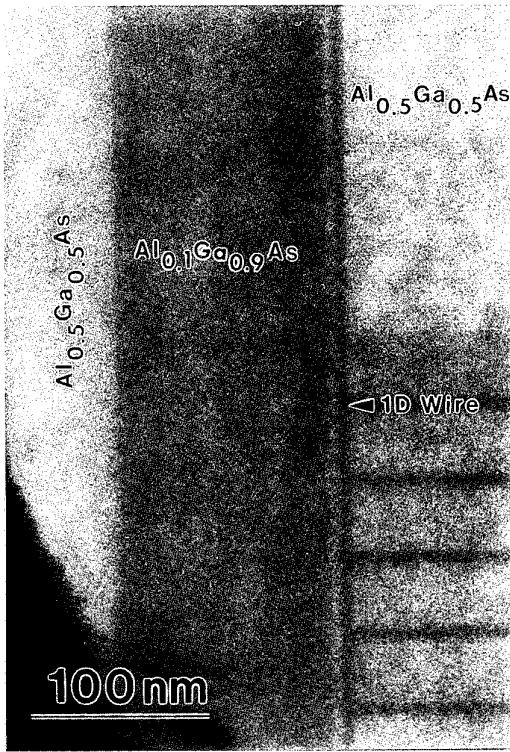


Figure 3. Cross-sectional bright-field transmission electron micrograph of the QWRs and the surrounding waveguide layers. The location of one T-shaped QW intersection is marked by an arrow.

electron micrograph in figure 3. The QWs appear in this image, taken with the electron beam aligned along the wire axis, as dark bands oriented parallel to the edges of the figure while the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layers show brighter contrast corresponding to their x values. It should be noted at this point that we could not detect any defects in the overgrown structure originating from the (110) cleave which served as a starting plane for the second MBE growth step.

Figure 4 compares emission spectra of our QWR laser below and above threshold for stimulated emission. At low excitation power (0.25 mW) exciton recombination in the QWRs and QWs is observed at about 1.563 and 1.58 eV respectively, indicating that the 1D state is about 17 meV deep with respect to the well. The doublet structure in the QW emission is believed to originate from slightly different confinement energies for the QWs grown along the [001] direction and the [110]-oriented QW formed during overgrowth. With increasing pump power the contrast in the Fabry-Pérot oscillations, which develop on the low-energy side of the QWR peak, increases, and at about 10 mW stimulated emission occurs. Further increase of the pump power leads to a significant narrowing of the QWR emission spectrum until at pump levels above about 30 mW laser operation takes place predominantly in a single longitudinal mode. The unusually strong contrast in the Fabry-Pérot oscillations even at the lowest excitation level indicates a high degree of transparency in the waveguide within this spectral

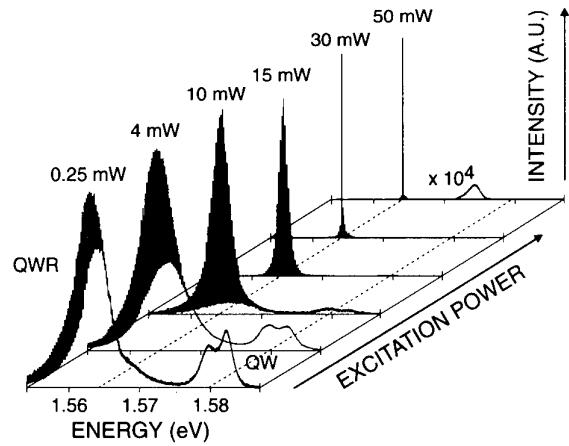


Figure 4. TE-polarized photoluminescence (PL) spectra recorded below, at and above threshold for stimulated emission in the QWRs ($P_{th} \sim 10$ mW). The orientation of the E vector for this polarization direction is given in figure 2.

region. This is due to the small fraction Γ of the optical intensity distribution, of about 3×10^{-3} , that overlaps with the QWRs. For comparison, the corresponding value for the QWs in our structure is $\Gamma \sim 0.15$. The laser output as a function of pump power is shown in figure 5 together with high-resolution spectra of the QWR emission. In contrast to the QW peak intensity, which increases linearly or sublinearly with excitation power, superlinear behaviour is observed. In addition, the QW photoluminescence (PL) signal red-shifts by as much as 5 meV with increasing

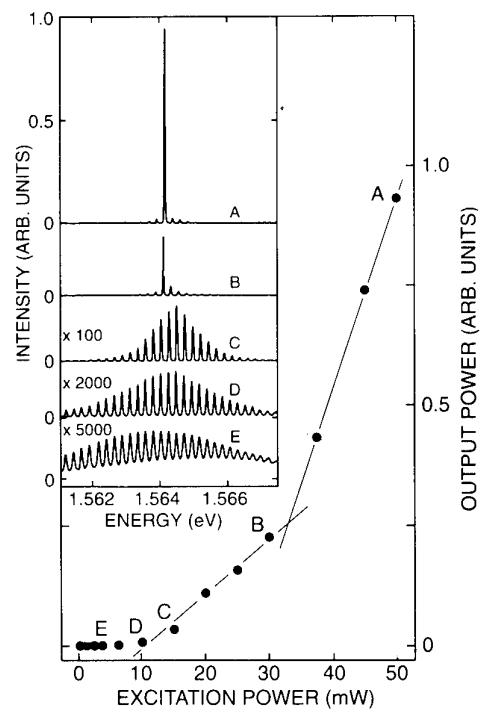


Figure 5. QWR laser output versus excitation power. High-resolution TE-polarized emission spectra recorded at 4, 10, 15, 30 and 50 mW excitation power are shown in the inset.

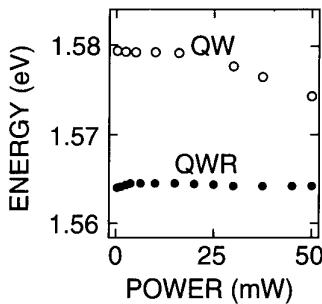


Figure 6. Power dependency of the QW and QWR photoluminescence energies.

pump power as depicted in figure 6. This is consistent with exciton ionization and photoexcitation of a free electron-hole plasma which is subject to bandgap renormalization effects. For 50 mW excitation ($\sim 3 \text{ kW cm}^{-2}$) we expect a sheet carrier density of $\sim 10^{11} \text{ electron-hole pairs/cm}^2$ assuming an absorption length of $0.5 \mu\text{m}$ in GaAs and a recombination time of 1 ns. The shrinkage of the free-particle bandgap for this density is about 15 meV [4] and exceeds the exciton binding energy in the QWS by about 5 meV. Strikingly, no appreciable shift in the QWR emission occurs for all pump powers used in this experiment. We interpret the implied absence of bandgap renormalization effects as the signature that electron-hole recombination in our QWRS is exclusively excitonic in character. To the best of our knowledge this represents the first observation of excitonic gain in a III-V semiconductor laser. It thus appears that the exciton gas phase in 1D is more stable against ionization than its 2D counterpart.

Excitonic gain has been recently reported for ZnSe QW lasers [20]. The enhanced stability of the excitons in this case is attributed to the large exciton binding energies in II-VI materials. In order to estimate the binding energies of the 1D excitons in our structure we have calculated the energetic difference in the free-carrier QW and QWR transitions. The results for electrons are plotted in figure 7 as a function of the QW width. In order to obtain the electron and hole wavefunctions for the T-shaped confining potential we have employed a two-dimensional transfer matrix technique using a one-band model for the

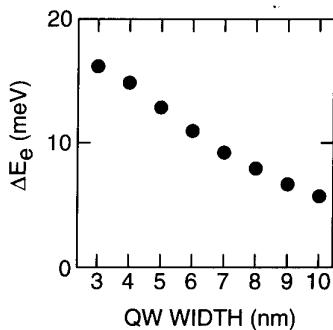


Figure 7. Calculated binding energy of electrons to the QWRS with respect to the lowest QW state as a function of QW width for equally thick wells.

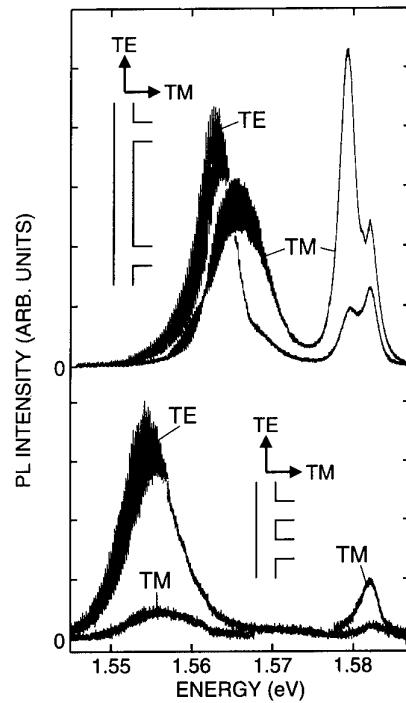


Figure 8. Comparison of low-power PL spectra (0.25 mW) of two QWR samples with centre-to-centre distances between T-intersections of 45 nm (upper panel) and 14 nm (lower panel).

hole with masses determined via the diagonal term in the Luttinger Hamiltonian with the angular momentum quantization axis parallel to the [110] overgrowth direction. In the absence of electron-hole Coulomb interaction, the holes are, owing to their larger mass, only weakly bound to the T-intersection. For 7 nm wide QWS the theoretically predicted value for the electron-heavy hole transition is 10 meV. This is considerably less than the observed red-shift of 17 meV. Since both the QWR and the QW transition at this low excitation level ($\sim 15 \text{ W cm}^{-2}$) are undoubtedly of excitonic nature, the energy difference of about 7 meV directly reflects the enhancement of the exciton binding energy due to the reduced dimensionality. Taking a 2D exciton binding energy of 10–11 meV [21, 22] for the 7 nm wide QWS into account, we obtain a 1D enhancement of more than 50% to about 17 meV for the QWRS. This considerably exceeds the largest previously reported 1D enhancement of 15% observed in lithographically defined structures with a cross section of about 70 by 14 nm [23].

Polarization-resolved PL spectra of the QWR laser sample recorded at low excitation power are displayed in the upper panel of figure 8. The QW response, which is dominated by the MQW emission, shows the typical polarization anisotropy of a 2D system. In contrast, the TE- (E parallel to the (110) cleavage plane) and TM-polarized (E parallel to the (001) substrate plane) QWR luminescence are of nearly equal intensity, although of different spectral shape and separated by about 3 meV. The TE-polarized QWR emission is characterized by a relatively narrow peak at 1.563 eV (full width at half

maximum ~ 4 meV). Shifted by about 5 meV, we observe a shoulder on the high-energy side of this peak indicating an additional transition with oscillator strength in this polarization direction. The TM-polarized PL signal originating from the QWRS is much wider and exhibits no additional structure. The large width of this peak of about 8 meV strongly suggests the presence of several TM-active transitions in this spectral region. We believe that these transitions occurring at higher energies arise from several closely spaced hole states in the valence band of our QWRS. An alternative interpretation would be the existence of excited states of the 1D exciton. However, this is unlikely in view of the specific symmetry of our QWR structure. We note that there exists a [001] mirror plane which contains the E vector in the TM polarization. For TE-polarized light the electric field is perpendicular to this mirror plane. Consequently, states of the same exciton series are not allowed in both TE and TM polarization.

Qualitatively, we can interpret the reduced polarization anisotropy as a signature of carrier confinement to 1D since the emission from a QWR with symmetrical cross section would be isotropic, neglecting the anisotropy of the crystal structure. In order to verify this point we have fabricated a sample in which the centre-to-centre distance between the T-shaped QW intersections has been reduced from 45 to 14 nm. In this case the electronic states at the T-junctions should be strongly coupled and indeed a QW-like polarization behaviour of the PL signal with a TE/TM intensity ratio of 7.7 is observed (lower panel of figure 8). However, to account quantitatively for the observed polarization behaviour, a detailed theoretical description of the band structure including effects caused by splitting and mixing of hole states in our particular wire geometry is required.

4. Conclusions and prospects

We have used CEO to fabricate nearly perfect QWRS with a cross section of only 7 by 7 nm. Stimulated excitonic emission in the 1D quantum limit, i.e. from the ground state exciton, was demonstrated at low effective pump power densities of 60 W cm^{-2} (power absorbed in the active region), where the QWS adjacent to the QWRS do not lase. Single-mode laser operation with no appreciable shift of the emission at all pump powers indicates an enhanced stability of the exciton gas phase in 1D and suggests interesting new behaviour for excitons in low-dimensional systems. As an extension of this work we point out that the intersection of three QWS results in the formation of a zero-dimensional (0D) bound state. A future implementation of the CEO method which involves two separate cleave and overgrowth steps should be well suited for the fabrication of a linear array of these 0D structures.

Acknowledgments

We appreciate valuable discussions with H L Stormer, N K Dutta, M S Hybertsen, A F J Levi and R E Slusher,

focused ion beam sample preparation by F Stevie and technical assistance from B S Dennis and D Bahnck. Werner Wegscheider gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft during the initial stages of this project.

References

- [1] Arakawa Y and Yariv A 1986 Quantum well lasers—gain, spectra, dynamics *IEEE J. Quantum Electron.* **22** 187–99
- [2] Asada M, Miyamoto Y and Suematsu Y 1986 Gain and the threshold of three-dimensional quantum-box lasers *IEEE J. Quantum Electron.* **22** 1915–21
- [3] Arakawa Y and Sakaki H 1982 Multidimensional quantum well laser and temperature dependence of its threshold current *Appl. Phys. Lett.* **40** 939–41
- [4] Schmitt-Rink S, Chemla D S and Miller D A B 1989 Linear and nonlinear optical properties of semiconductor quantum wells *Adv. Phys.* **38** 89–188
- [5] Degani M H and Hipolito O 1987 Exciton binding energy in quantum-well wires *Phys. Rev. B* **35** 9345–8
- [6] Bányai L, Galbraith I, Ell C and Haug H 1987 Excitons and biexcitons in semiconductor quantum wires. *Phys. Rev. B* **36** 6099–104.
- [7] Ogawa T and Takahara T 1991 Interband absorption spectra and Sommerfeld factors of a one-dimensional electron–hole system *Phys. Rev. B* **43** 14325–8; 1991 Optical absorption and Sommerfeld factors of one-dimensional semiconductors: an exact treatment of excitonic effects *Phys. Rev. B* **44** 8138–56
- [8] Ivanov A L and Haug H 1993 Existence of exciton crystals in quantum wires *Phys. Rev. Lett.* **71** 3182–5
- [9] For review see C Weisbuch and B Vinter 1991 *Quantum Semiconductor Structures* (San Diego: Academic)
- [10] Kapon E, Hwang D M and Bhat R 1989 Stimulated emission in semiconductor quantum wire heterostructures *Phys. Rev. Lett.* **63** 430–3
- [11] Tsuchiya M, Gaines J M, Yan R H, Simes R J, Holtz P O, Coldren L A and Petroff P M 1989 Optical anisotropy in a quantum-well-wire array with two-dimensional quantum confinement *Phys. Rev. Lett.* **62** 466–9
- [12] Nötzel, R, Ledentsov N N, Däweritz L, Hohenstein M and Ploog K 1991 Direct synthesis of corrugated superlattices on non-(100)-oriented substrates *Phys. Rev. Lett.* **67** 3812–15
- [13] Tsukamoto S, Nagamune Y, Nishioka M and Arakawa Y 1993 Fabrication of GaAs quantum wires (~ 10 nm) by metalorganic chemical vapor selective deposition growth *Appl. Phys. Lett.* **63** 355–7
- [14] Brunner K, Bockelmann U, Abstreiter G, Walther M, Böhm G, Tränkle G and Weimann G 1992 Photoluminescence from a single GaAs/AlGaAs quantum dot *Phys. Rev. Lett.* **69** 3216–19
- [15] Prins F E, Lehr G, Burkard M, Schweizer H, Pillkuhn M H and Smith G W 1993 Photoluminescence excitation spectroscopy on intermixed GaAs/AlGaAs quantum wires *Appl. Phys. Lett.* **62** 1365–7
- [16] Wegscheider W, Pfeiffer L N, Dignam M M, Pinczuk A, West K W, McCall S L and Hull R 1993 Lasing from excitons in quantum wires *Phys. Rev. Lett.* **71** 4071–4
- [17] Pfeiffer L, Stormer H L, Ashoori R C, Goñi A R, Pinczuk A, Baldwin K W and West K W 1992 *Low-Dimensional Electronic Systems* ed G Bauer, F Kuchar and H Heinrich (Berlin: Springer)

- [18] Pfeiffer L, Störmer H L, Baldwin K W, West K W, Goñi A R, Pinzuk A, Ashoori R C, Dignam M M and Wegscheider W 1993 Cleaved edge overgrowth for quantum wire fabrication *J. Crystal Growth* **127** 849–57
- [19] Goñi A R, Pfeiffer L N, West K W, Pinckuk A, Baranger H U and Stormer H L 1992 Observation of quantum wire formation at intersecting quantum wells *Appl. Phys. Lett.* **61** 1956–8
- [20] Ding J, Jeon H, Ishihara T, Hagerott M and Nurmikko A V 1992 Excitonic gain and laser emission in ZnSe-based quantum wells *Phys. Rev. Lett.* **69** 1707–10
- [21] Greene R L, Bajaj K K and Phelps D E 1984 Energy levels of Wannier excitons in Ga-Ga_{1-x}Al_xAs quantum-well structures *Phys. Rev. B* **29** 1807–12
- [22] Andreani L C and Pasquarello A 1990 Accurate theory of excitons in GaAs-Ga_{1-x}Al_xAs quantum wells *Phys. Rev. B* **42** 8928–38
- [23] Kohl M, Heitmann D, Grambow P and Ploog K 1989 One-dimensional magnetooexcitons in GaAs/Al_xGa_{1-x}As quantum wires *Phys. Rev. Lett.* **63** 2124–7

Recombination kinetics and intersubband relaxation in semiconductor quantum wires

M Grundmann[†], J Christen^{†‡}, M Joschko[†], O Stier[†], D Bimberg[†]
and E Kapon[§]

[†] Institut für Festkörperphysik, TU Berlin, Hardenbergstr. 36, D-10623 Berlin, Germany

[§] Institute of Micro- and Optoelectronics, Swiss Federal Institute of Technology,
CH-1015 Lausanne, Switzerland

Abstract. We present a detailed and systematic investigation of carrier capture, relaxation, cooling and radiative recombination in a one-dimensional semiconductor quantum wire of high structural perfection and optical quality over a large range of excitation (carrier) densities. Experimental evidence for a complete lack of 1D bandgap renormalization is found. Even up to high carrier densities, $>10^6 \text{ cm}^{-1}$, where strong band filling is already present and directly visible in the luminescence, no shift of bandgap to low energy is found. The carrier cooling in 1D is appreciably slower than in comparable 2D structures, thus leading to high carrier temperatures. This confirms theoretical predictions of reduced phonon scattering probability in one-dimensional structures. The temperature dependence of the radiative lifetime of the 1D carriers is investigated. The theoretically predicted \sqrt{T} dependence is not found. On the contrary an empirical $\tau_{\text{rad}} = 0.02 T \text{ ns K}^{-1}$ law is fulfilled.

1. Introduction

Physics in one-dimensional materials is fundamentally different from physics in two- or three-dimensional materials. A number of challenging theoretical predictions have been made which remain to a large part unconfirmed. Experimental verification so far has been hampered due to the lack of structurally perfect quantum wires with high optical quality. The confinement in two dimensions results in a 1D subband structure together with a modified density of states (dos) which shows a $1/\sqrt{E}$ singularity at each 1D subband edge. The disappearance of continuous dispersion in two dimensions hinders energy relaxation because energy and k conservation are more difficult to match simultaneously. Thus reduced energy relaxation rates [1] and subsequently higher carrier temperatures are predicted. The radiative lifetime, which shows a linear dependence on the temperature T for quantum wells (QWLS), is predicted to exhibit a \sqrt{T} dependence for quantum wires (QWRs) as a direct consequence of the changed dos [2].

In this paper we will present results of a detailed investigation of carrier relaxation, cooling and recombination in semiconductor quantum wires across a large range of excitation density. Apart from these intrinsic properties of the true 1D carriers, we will address the capture of carriers into the QWR through adjacent QWLS which often dominates the experimentally observed decay.

2. Experiment

In the following we will briefly describe sample preparation by OMVPE and our luminescence characterization methods. Our approach to the fabrication of true quantum wires by OMVPE (organometallic vapour-phase epitaxy) of quantum wells on patterned substrates has already been described in detail in [3]. We would like to emphasize here only the key points and the sample-specific parameters. Following wet chemical etching of V-grooves with $3.5 \mu\text{m}$ pitch into a GaAs substrate, a $0.7 \mu\text{m}$ Al_{0.45}Ga_{0.55}As layer has been grown. The growth kinetics of the AlGaAs lead to a self-organized sharpening of the V-groove to a minimal radius of curvature (at the lower interface of the QWR) of about 12 nm intrinsically defined by the growth conditions. Then a 3 nm GaAs QWL is deposited; in the bottom of the groove a crescent-shaped QWR with a maximal thickness of 8 nm in the centre of the groove is formed. The cross-sectional area of the resulting QWR is about 250 nm^2 . Subsequently another 50 nm of AlGaAs barrier and a 20 nm GaAs cap are grown on top of the GaAs layer. A transmission electron microscope image of the central part of the groove is shown in figure 1. In the centre of the groove a dark vertical line is visible which is due to the so-called vertical quantum well [3]. The preferential Ga diffusion into the groove, responsible for the formation of the QWR, also takes place in the AlGaAs, resulting in a higher Ga content in the centre. The various QWLS and the QWR will manifest themselves later on in the luminescence spectra at their particular recombination energies.

[‡] Permanent address: Otto-von-Guericke Universität, D-39016 Magdeburg, Germany.

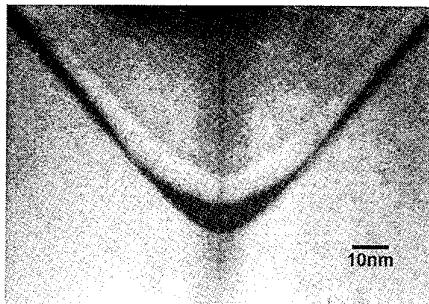


Figure 1. TEM image of a single GaAs/AlGaAs QWR structure.

For luminescence characterization of the structure we employed spectrally, laterally and time-resolved cathodoluminescence (CL). In the CL experiment it is possible to investigate properties either *laterally averaged* over a large sample area or *locally* with a high lateral resolution of better than 60 nm [4]. In this way the specific properties of an isolated specific *single* QWR are investigated. As the detector for non-time-resolved measurements we use an infrared intensified reticon. By using electrostatic beam blanking we perform transient experiments with an overall time resolution of 30 ps, limited by the MCP photomultiplier (single-photon counting technique). Details about the CL system and the different data collection modes are given in [4, 5].

3. Lateral bandgap modulation of GaAs QWL

The GaAs QWL on the corrugated AlGaAs shows different layer thickness depending on which plane of the structure it is grown on. The laterally averaged plane view CL spectrum at low sample temperature, $T = 5$ K, is shown in figure 2. A variety of transitions are observed which are due to carrier recombination in the AlGaAs barrier, the vertical QWL in the centre of the groove, the sidewall QWL, the top QWL (showing monolayer splitting) and in the QWR itself. The QWR luminescence is weak due to the small volume filling factor ($\approx 10^{-3}$) of the QWR as compared with the whole structure. To directly visualize the lateral modulation of bandgap a lateral CL linescan across the wire is displayed in figure 3. In figure 3(a) the secondary electron (SE) image of the sample in plane view is shown. In the upper and lower part of the figure the top ridge is visible, while in the middle the sidewalls with the QWR in the centre show up. In figure 3(b) the CL spectrum along a linescan perpendicular to the wire (as indicated by the white line in figure 3(a)) is displayed. The x-axis is now given by the wavelength of emission while the y-axis is the lateral position along the linescan. The CL intensity is given on a logarithmic scale to display the full dynamic range. The top QWL shows almost no variation in bandgap energy; only directly at the edge close to the sidewall does a second peak at lower energy appear indicating a thicker region there. The sidewall QWL exhibits a recombination wavelength of 700 nm at the edge to the top QWL which gradually increases to

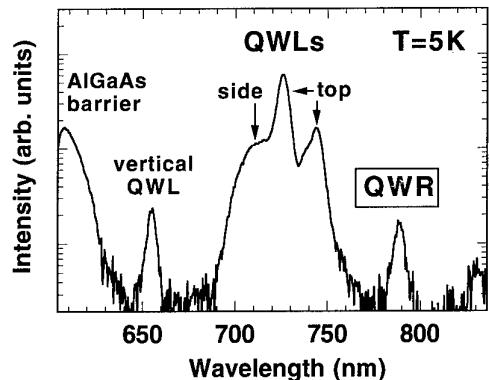


Figure 2. Laterally averaged CL spectrum at low temperature, $T = 5$ K. The area of excitation is $200 \times 300 \mu\text{m}^2$ (in plane view).

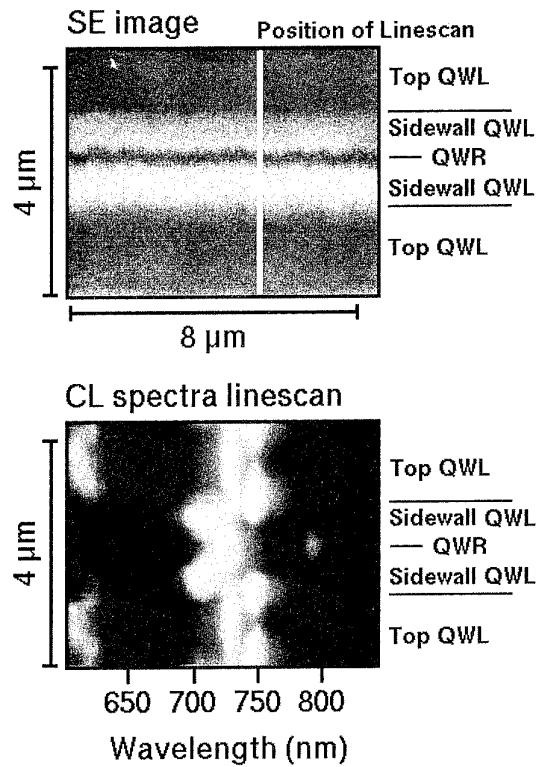


Figure 3. Plane view SE image (upper part) of QWR sample, showing top and sidewall with the QWR in the centre. The white line indicates the position of the linescan on which the CL spectra linescan (lower part) has been taken. The CL intensity is given on a logarithmic grey scale.

about 730 nm right at the centre of the V-groove. This directly visualizes the linear tapering of the sidewall QWL from about $L_z = 2.1$ nm at the edge to $L_z = 3$ nm in the centre.

4. Carrier capture

After fast capture from the barrier into the QWLs and QWR [6], excess carriers will diffuse into the QWR via the adjacent sidewall quantum well. Its tapering induces an

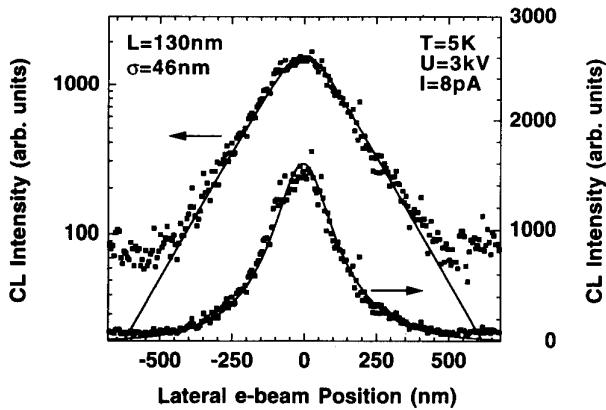


Figure 4. Lateral profile of the QWR intensity in logarithmic and linear scale for lateral variation of the electron beam excitation spot. Excitation conditions were $U = 3$ kV, $I = 8$ pA at $T = 5$ K. The width of the figure is as long as the side quantum well. The full curves are theoretical fits with a diffusion profile (diffusion length $L = 130$ nm) convoluted with a finite experimental resolution (Gaussian with $\sigma = 46$ nm).

additional drift current. The virtual lateral extension with which the wire appears in the luminescence linescan is given by three quantities: the real (geometrical) wire width, the carrier diffusion and the finite resolution of the CL experiment as given by the generation volume of excess carriers. The profile of the QWR intensity across the groove is shown in figure 4. We fit the data with an $\exp(-|x|/L)$ diffusion profile (L being the diffusion length) convoluted with a Gaussian due to the finite resolution of the CL experiment. In a conservative estimate we neglect the finite geometrical size of the wire. Assuming a diffusion length of $L = 130$ nm a close to perfect fit is obtained. The Gaussian broadening is consistently fitted to $\sigma = 46$ nm, providing the lateral resolution of our CL experiment [4].

5. Formation of 1D subband structure

The confinement in two dimensions is expected to lead to one-dimensional subbands. The subband spacing largely depends on the depth of the potential well and the wire size and shape. Since our wires have no simple cross-sectional shape (like rectangular or circular) we have to solve the two-dimensional Schrödinger equation in the actual potential to obtain the correct confinement energies and the proper wavefunctions. This approach goes beyond the adiabatic approximation [7] where the problem is separated into solving the Schrödinger equation first for the (stronger) confinement in growth direction and subsequently in the (weaker) lateral confinement potential. The calculated eigenfunctions are depicted in figure 5 for the first three electron levels together with the expectation values of the spatial extension in the x and z directions. As can be easily seen, the nodes appear in the lateral direction, indicating that all subbands are due to the lateral confinement in

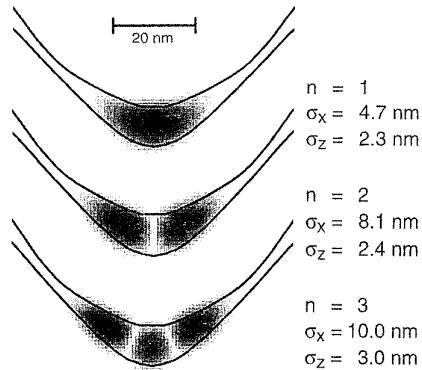


Figure 5. Electron wavefunction for the first three confined levels (Ψ^2 in a logarithmic grey scale) together with the expectation values for the extension in x (horizontal) and z (vertical) directions. σ_x is defined as $\sigma_x = \langle x^2 - \langle x \rangle^2 \rangle^{0.5}$, $\sigma_z = \langle z^2 - \langle z \rangle^2 \rangle^{0.5}$.

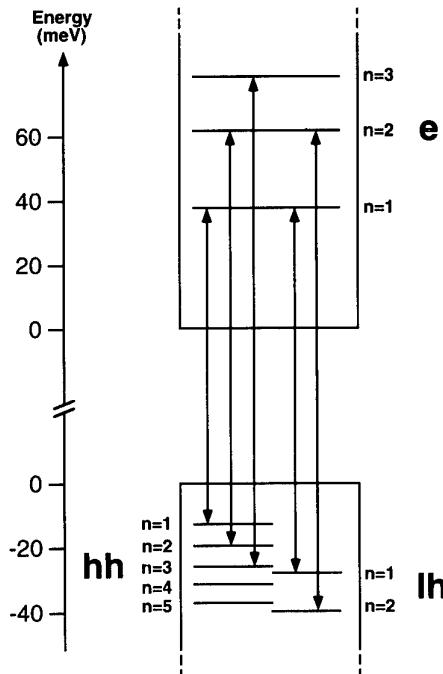


Figure 6. Energy diagram for the lowest confinement levels in the single QWR from figure 1. The energetically lowest intersubband transitions are indicated with arrows.

agreement with the adiabatic approximation. Similar wavefunctions are obtained for the holes. The calculated energy diagram is shown in figure 6; the electrons exhibit subband separation of about 20 meV, while the heavy-hole levels are separated by about 6 meV. The first three heavy-hole levels show a smaller energy shift to the lowest confined light-hole level. (When obtaining the spectrum we note that, due to the mirror symmetry of the QWR with respect to the groove centre, parity is still a good quantum number.)

The calculated 1D intersubband transitions are indicated in figure 7 together with a CL spectrum

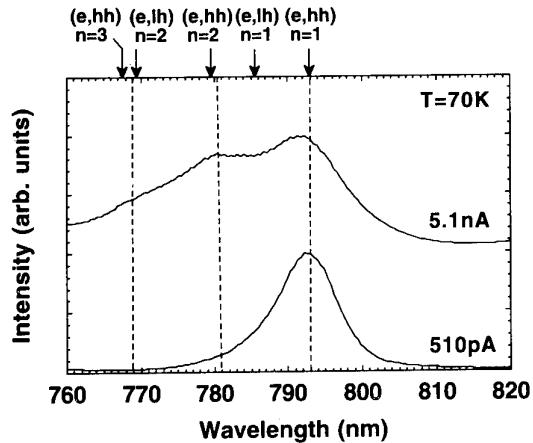


Figure 7. CL spectrum of the QWR at $T = 70$ K for two different excitation densities, corresponding to carrier densities of $1 \times 10^5 \text{ cm}^{-1}$ and $1 \times 10^6 \text{ cm}^{-1}$. The theoretical peak positions of the lowest-lying transitions are indicated by arrows.

at $T = 70$ K (where higher subbands are sufficiently populated) for two different excitation densities, corresponding to carrier densities of $1 \times 10^5 \text{ cm}^{-1}$ and $1 \times 10^6 \text{ cm}^{-1}$. Beside the (e, hh), $n = 1$ groundstate transition at 793 nm with FWHM = 13 meV two higher transitions can be recognized. The first higher transition at 781 nm agrees very well with the assignment (e, hh), $n = 2$. The (e, lh), $n = 1$ transition, which is expected at lower transition energy, does not show up in the spectrum with a distinct peak. Since the hh, $n = 2$ level has a smaller confinement energy than the lh, $n = 1$ it is obvious that it has a higher population (see figure 6). The e, $n = 2$ band, however, is filled due to band filling of the e, $n = 1$ band for the high excitation density used in figure 7. The appearance of the third transition at 769 nm, which involves the e, $n = 2$ and/or e, $n = 3$ level, also indicates high population of the higher electron subband(s), which is consistent with theoretical calculations of the chemical potential μ_e .

In figure 8 the intensities at the $n = 1$ (793 nm) and $n = 2$ (781 nm) transitions are plotted as a function of beam current, i.e. the excitation density. The ground state starts to saturate at a beam current of about $I = 1$ nA which corresponds to a carrier concentration of $2 \times 10^5 \text{ cm}^{-1}$ in the QWR. At $n \sim 10^6 \text{ cm}^{-1}$ all states in the e, $n = 1$ subband up to the e, $n = 2$ subband edge are filled. Beginning with the saturation of the groundstate, the second transition at 781 nm starts to appear.

The excitation dependence of the bandgap for lattice temperatures $T = 5, 40, 70$ and 100 K is depicted in figure 9. The peak position of the ground-state shows, for the highest excitation densities only, a minor shift to higher energy. Also the $n = 2$ level remains at a constant energy position. We note that this is direct evidence for the complete absence of bandgap renormalization effects, i.e. shift of the bandgap towards lower recombination energy with increasing excitation density as observed in 2D systems [8]. This is possibly due to enhanced excitonic and correlation effects, demanding more detailed theoretical investigation including lineshape-fitting.

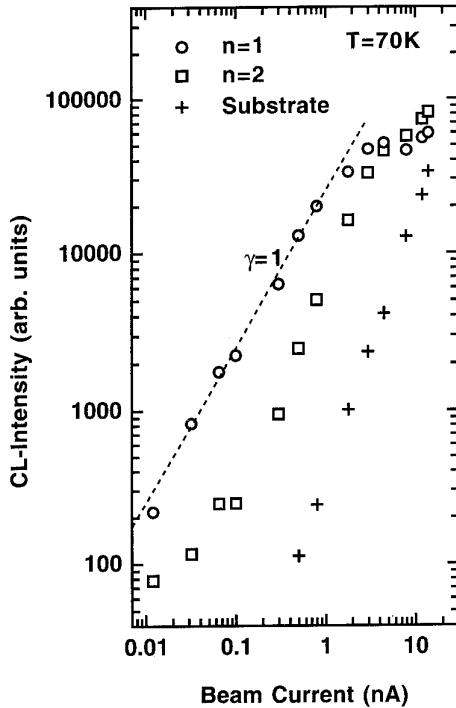


Figure 8. Intensity at 793 nm (ground state, $n = 1$) and 781 nm ($n = 2$ level) as a function of excitation density, varying by three orders of magnitude.

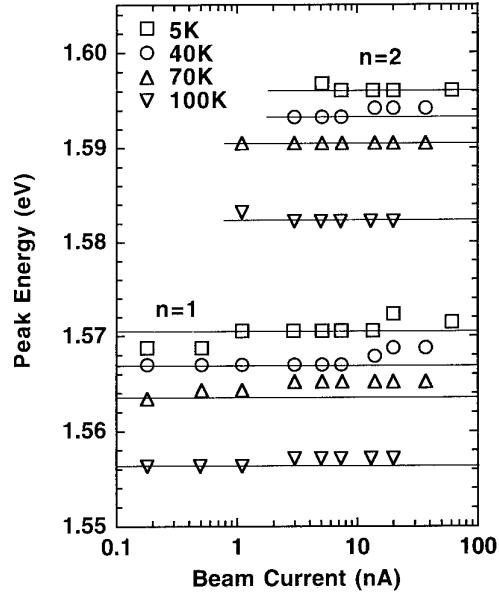


Figure 9. Energetic position at lattice temperatures $T = 5, 40, 70$ and 100 K of the $n = 1$ and $n = 2$ transitions as function of the excitation density. For beam currents $I > 1$ nA the ground state shows strong filling and saturation. Horizontal lines are peak positions at the lowest excitation density.

6. Carrier relaxation and cooling in 1D

In this section we describe the intersubband relaxation of the 1D charge carriers and their cooling. The sample is excited with rectangular electron pulses which are

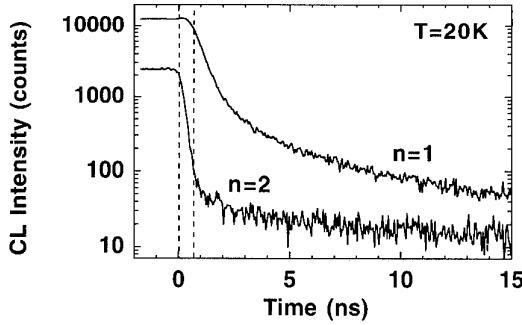


Figure 10. CL transients at $T = 20\text{ K}$, $I = 3\text{ nA}$ on the $n = 1$ and $n = 2$ levels. The left broken vertical line indicates the switch-off time of the electron pulse; the right broken vertical line indicates the time when $n = 2$ is almost completely relaxed and $n = 1$ starts to show a single exponential decay.

sufficiently long (about 40 ns) to reach quasi equilibrium in the sample. When the electron beam is switched off, the transient is initiated. We point out that this approach is rather different from time-resolved photoluminescence where usually a picosecond pulse is applied and the electronic system is not in quasi equilibrium when the decay starts [9]. In figure 10 we show the transients on the $n = 1$ and the $n = 2$ levels at $T = 20\text{ K}$ and a moderate excitation density so that the second level is sufficiently populated. The $n = 2$ luminescence shows a perfect single exponential decay over two orders of magnitude with a time constant of 200 ps. In contrast, after the excitation source has been switched off, the intensity on the $n = 1$ peak remains constant for some 400 ps. Then an exponential decay with a time constant of $\tau_{\text{total}} = 500\text{ ps}$ is observed. The $n = 1$ level is obviously fed from the $n = 2$ level at the beginning of the transient. Obviously the $n = 2$ level has two loss channels: its own radiative decay and the transfer into the $n = 1$ level. The exponential decay constant of the $n = 1$ level is given by the carrier lifetime of the groundstate. For a quantum efficiency η of unity (i.e. the recombination is dominated by radiative processes at low temperature, see next section) this directly gives the radiative recombination time constant τ_{rad} . Even in the case of $\eta \neq 1$, the radiative lifetime is still proportional to the ratio of the total lifetime and the intensity because $\tau_{\text{rad}} = \tau_{\text{total}}/\eta$.

During the relaxation process the charge carriers cool, owing to phonon scattering. Since the subband separations are $\leq 21\text{ meV}$ and nearly all the spectral emission of the $n = 1$ and $n = 2$ state is within 30 meV of the fundamental bandgap of the QWR Fröhlich interaction is efficiently suppressed and the main energy relaxation mechanism will be acoustic phonon scattering of the holes [9].

For quantifying the cooling rate we measured time-delayed (TD) spectra during the transient which directly show the decrease of carrier temperature with time. In figure 11 TD spectra for the transients at a lattice temperature $T_L = 20\text{ K}$ of figure 10 are shown during the pulse (steady state) and at different delay times after the

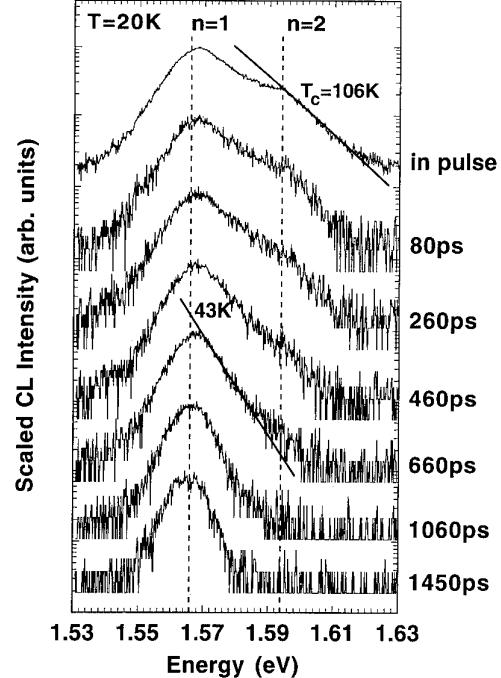


Figure 11. Time-delayed spectra of the QWR luminescence at $T = 20\text{ K}$. The time windows correspond to the time axis from figure 10. Two exponential slopes fitted to the high-energy tail of the spectrum with their respective carrier temperature are included.

excitation is switched off. The exponential slope of the high-energy tail of the spectrum becomes steeper with increasing delay time, directly reflecting the cooling of the 1D carrier gas. Additionally a faster relaxation of the $n = 2$ level as compared with the $n = 1$ level is obvious.

The carrier temperature T_c is obtained from fitting an exponential slope $\exp(-E/kT_c)$ into the high-energy tail of the spectrum. This procedure, which will also be applied below to obtain the cooling rate from time-delayed spectra, is experimentally justified by the actual observation of an exponential tail over several orders of magnitude. However, a full lineshape analysis should be made taking into account the shape of the joint density of states with bound and ionized excitonic states. The Coulomb interaction is expected to field a Sommerfeld factor removing the $1/\sqrt{E}$ singularity in the optical DOS at the band edge resulting in a sufficiently smooth DOS. Thus the true carrier temperature T_c might be slightly different from what is reported here, as a result of a more detailed analysis.

As long as the $n = 2$ level is populated we obtain the carrier temperature from the slope of the $n = 2$ tail; for delay times longer than 400 ps we determine the carrier temperature from the slope of the $n = 1$ transition, presuming an equal carrier temperature for the $n = 2$ and $n = 1$ subbands due to efficient (but elastic) carrier-carrier scattering. The carrier temperature as a function of time is plotted in figure 12 (the broken curve is a guide to the eye). After 700 ps the cooling process slows down appreciably and the carrier temperature remains almost

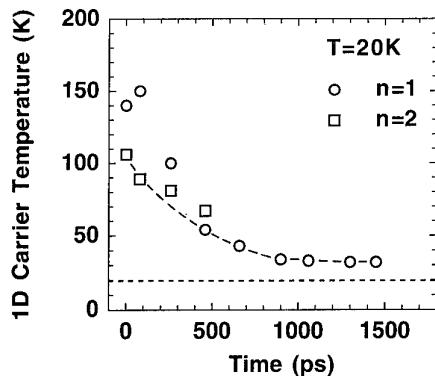


Figure 12. 1D carrier temperature T_c as obtained for the $n = 1$ and $n = 2$ levels from the TD spectra of figure 11 as a function of delay time after the excitation pulse. The valid T_c for each time is the lower value; the broken curve is a guide to the eye.

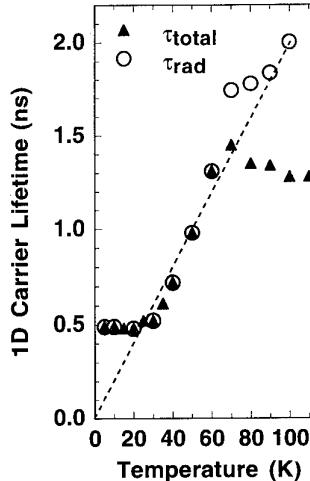


Figure 13. Total lifetime and radiative lifetime of the ground state QWR luminescence as a function of lattice temperature.

constant at a value of $T_c = 30$ K for $T_L = 20$ K. Estimating an average energy relaxation rate for the first 900 ps yields $\langle dE/dt \rangle_{av} = 3.8$ meV ns $^{-1}$. This value is about 50% less than the value obtained by Polland *et al* [10] for cooling by solely acoustic deformation potential scattering in an $L_z = 5$ nm GaAs/Al_{0.30}Ga_{0.70}As QWL. A detailed discussion will be given elsewhere [11]. The slow energy relaxation measured by us is also consistent with the results of Mayer *et al* [12] obtained from intermixed GaAs/AlGaAs QWLs, although their structures show smaller subband spacing (10 meV for electrons) and larger inhomogeneous broadening (FWHM = 22 meV). We attribute the slow energy relaxation to the reduced phonon emission probability in the 1D system as predicted by theory ('phonon bottleneck' effect). For wires with dimensions less than 10 nm \times 40 nm in particular, pronounced effects are predicted [1].

7. Radiative carrier recombination in quantum wires: 1D lifetimes

In spatially and time-resolved CL experiments, i.e. generating 1D carriers locally directly in the QWR, we can separate the carrier capture from the real 1D carrier recombination and we are able to measure the true 1D carrier lifetimes. The radiative recombination time constant is a function of the temperature: due to k conservation only those excitons with centre-of-mass k -vector close to zero can recombine leading to a theoretically predicted $\tau_{rad} \propto \sqrt{T}$ dependence [2] via thermalization within the 1D DOS. However, the measured dependence looks very different from the prediction. In figure 13 the total recombination time constant as obtained from fitting an exponential decay $\exp(-t/\tau_{total})$ into the transients is plotted together with the radiative lifetime $\tau_{rad} \propto \tau_{total}/\text{intensity}$ as discussed above. At lowest temperatures the lifetime is almost constant, indicating that the spatial coherence of the exciton is limited by a localization mechanism [13]. With

increasing temperature we observe a *linear* increase $\tau_{rad} = 0.02T$ ns K $^{-1}$ in complete contrast to the theoretical prediction. Akiyama *et al* [14] made a similar experiment on MBE-grown QWRS of about the same size and claimed to observe the predicted \sqrt{T} dependence of the total lifetime; however, re-evaluation of their data, taking into account the measured drop of quantum efficiency with increasing temperature, reveals a *linear* increase of the radiative lifetime following $\tau_{rad} = 0.019T$ ns K $^{-1}$, which is identical to our results. Thus the only two systematic experimental investigations of the temperature dependence of the radiative lifetime in 1D reported so far are obviously in complete contrast to the theoretical prediction.

One cause for the discrepancy between theory and experiment is that the lifetime depends on the *carrier* temperature but has been plotted as a function of the *lattice* temperature. From our experiments about carrier cooling we know that due to slower energy relaxation in 1D systems the carrier temperature is appreciably larger than the lattice temperature. An improved theoretical treatment thus has to take into account cooling and radiative recombination *simultaneously* in a more general approach.

8. Conclusion

We have presented a detailed investigation of carrier relaxation, cooling and recombination in a one-dimensional semiconductor system. We have evidence for the complete absence of bandgap renormalization effects up to carrier densities of several 10⁶ cm $^{-1}$ where strong band filling is observed. This could be the indication of strong excitonic and carrier correlation effects as a particular effect of the reduced dimensionality of the carriers. Carrier cooling is found to be considerably slower than in comparable 2D systems. This is attributed to reduced phonon emission probability as a direct

consequence of the 1D density of states. Enhanced screening of the carrier–phonon interaction may also play an important role. The temperature dependence of the radiative recombination lifetime does not show the theoretically predicted \sqrt{T} dependence but follows an empirical $\tau_{\text{rad}} = 0.02 T \text{ ns K}^{-1}$ law. The reason for this discrepancy is probably the high 1D carrier temperature, which is always larger than lattice temperature. Improved theoretical modelling therefore has to take into account the energy relaxation and the radiative recombination at $K \approx 0$ simultaneously.

Acknowledgment

We are grateful to E Colas and R Bhat for supplying the samples and to D M Hwang for TEM analysis.

References

- [1] Bockelmann U and Bastard G 1990 *Phys. Rev.* **42** 8947
- [2] Citrin D S 1992 *Phys. Rev. Lett.* **69** 3393
- [3] Bhat R, Kapon E, Hwang D M, Koza M A and Yun C P 1988 *J. Crystal Growth* **93** 850
- [4] Christen J, Grundmann M and Bimberg D 1994 *Progress in Crystal Growth and Characterization of Materials* ed B Mullin (Oxford: Pergamon) at press
- [5] Christen J, Grundmann M and Bimberg D 1991 *J. Vac. Sci. Technol. B* **9** 2358
- [6] Christen J, Grundmann M, Kapon E, Colas E, Hwang D M and Bimberg D 1992 *Appl. Phys. Lett.* **61** 67
- [7] Kapon E, Hwang D M and Bhat R 1989 *Phys. Rev. Lett.* **63** 430
- [8] Tränkle G, Leier H, Forchel A, Haug H, Ell C and Weimann G 1987 *Phys. Rev. Lett.* **58** 419
- [9] Bimberg D, Müntzel H, Steckenborn A and Christen J 1985 *Phys. Rev. B* **31** 7788
- [10] Polland H-J, Rühle W W, Kuhl J, Plcog K, Fujiwara K and Nakayama T 1987 *Phys. Rev. B* **35** 8273
- [11] Christen J, Grundmann M, Bimberg D and Kapon E 1994 *Proc. ICPS-22, Vancouver* at press
- [12] Mayer G, Prins F E, Lehr G, Schweizer H, Leier H, Maile B E, Straka J, Forchel A and Weimann G 1993 *Phys. Rev. B* **47** 4060
- [13] Feldmann J, Peter G, Göbel E O, Dawson P, Moore K, Foxon C and Elliot R J 1987 *Phys. Rev. Lett.* **59** 2377
- [14] Akiyama H, Koshiba S, Someya T, Wada K, Noge H, Nakamura Y, Inoshita T, Shimizu A and Sakaki H 1994 *Phys. Rev. Lett.* **72** 924

Spectroscopy of field-effect-induced superlattices

W Hansen[†], A Schmeller[†], H Drexler[†], J P Kotthaus[†], M Holland[‡], G Tränkle[§], G Böhm[§] and G Weimann[§]

[†] Sektion Physik, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany

[‡] Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, UK

[§] Walter Schottky Institut, Am Coulombwall, 85748 Garching, Germany

Abstract. Interdigitated gates are employed on heterostructure surfaces to induce strong and tunable lateral superlattice potentials of type II, i.e. of purely field-effect origin. The design of the heterostructures is optimized in a way that avoids statistical potential fluctuations even at very low electron densities where screening of carriers is ineffective. Single-particle electron states confined to one-dimensional electron wires are investigated with capacitance spectroscopy. The precise control of the potential as well as the homogeneity of the potential allow us to investigate 'true' quantum wire arrays, i.e. wires in which only the lowest one-dimensional subband is occupied. Furthermore, we employ interdigitated gates to investigate the optical properties of quantum wells in a type II potential superlattice. The lateral fields are sufficiently high to field-ionize optically excited electron-hole pairs. Absorption spectra of the quantum well under such strong lateral fields show the typical signatures of the Franz-Keldysh effect.

1. Introduction

Precisely controllable lateral potentials are of wide interest for the study of one- and zero-dimensional electron systems. They allow the independent control of the potential amplitude as well as the density of electrons that occupy the electronic states in the superlattice. We fabricate interdigitated Schottky gates on the surface of heterostructure crystals to induce such tunable lateral type II potential superlattices with purely field-induced carrier systems. In contrast most of the previous work has been performed on modulation-doped heterojunctions containing two-dimensional electron systems at zero gate voltage. Although this approach is very successfully employed to investigate single electron dots (McEuen *et al* 1991, Ashoori *et al* 1993) or short electron wires (van Wees *et al* 1988, Wharam *et al* 1988), it suffers from inherent limitations. The Coulomb fields of the dopants in the barrier inflict statistical fluctuations upon the lateral potential superlattice which become very important if narrow electron wires or small dots are generated. In order to form the wires or dots the electron system is depleted beneath the Schottky gate with negative gate voltages applied between the front gate and the electron system. The Coulomb fields of the dopants in the barrier become increasingly effective at small electron densities, where screening of the electron system becomes less effective (Nixon and Davies 1990). Since the gate voltage is applied with respect to the electron system itself, the generation of a confinement potential is inevitably

associated with a reduced electron density and correspondingly a lateral geometry that is severely spoiled by the Coulomb fields of the dopants. This fact is especially disadvantageous if a precise control of the effective confinement potential is desired, even at very low densities and strong confinement potentials, where, for example, a quantum wire is in the one-dimensional quantum limit.

We present here a technique to avoid this problem. We induce lateral superlattice potentials with an interdigitated gate pattern fabricated with electron beam lithography on the surface of a heterojunction crystal. The crystal contains no dopants in the front barrier. For investigations of one-dimensional electron wires we supply a back electrode buried in the heterojunction substrate. Electrons can be injected from this back electrode through a shallow tunnel barrier into the quantum wires or quantum dots residing at the heterojunction interface.

Our experiments demonstrate that interdigitated gates can also be very successfully applied to study the impact of a lateral superlattice potential on the optical properties of quantum wells. In conventional heterostructures charge accumulation and leakage currents in the barrier region make such investigations at least difficult. The effect of a type II superlattice on the absorption or emission properties at bandgap energies is quite involved. Here the interaction of the electrons with photoexcited holes comes into play. Since an exciton is a neutral particle it would not feel the potential

superlattice if it was not polarizable. Because of its polarizability it is accelerated into the regions of maximum field. In sufficiently strong lateral fields the exciton dissociates. Subsequently, the electrons and holes accumulate in their respective potential minima and can considerably screen the potential superlattice until further charge accumulation is terminated by spatially direct or indirect electron-hole recombination. The fact that the processes are quite complex and depend on additional properties of the experimental system as well may be the reason why at present no common understanding of luminescence and absorption properties of quantum wells with potential superlattices seems to exist. Furthermore, luminescence and absorption properties are currently investigated mostly in shallow- or deep-etched quantum well samples. At least in the case of shallow-etched devices it is generally believed that the effective potential originates from charges on the etched surface, and thus is also a purely electrostatically induced type II potential. In these structures the potential modulation is hardly controllable. The role of stress patterns (Kash *et al* 1991) in shallow- and deep-etched heterojunctions originating from the corrugated crystal surface as well as the influence of process-induced damage are at present unresolved. In contrast, with our samples we can generate a tunable pure type II superlattice as is demonstrated by the observation of the lateral Franz-Keldysh (Franz 1958, Keldysh 1958) effect at voltages as low as $V_d \approx 1$ V between the fingerpairs of the interdigitated gate.

2. Interdigitated gates

The geometry of an interdigitated gate is sketched in figure 1. The electrodes form two interlocked fingergates

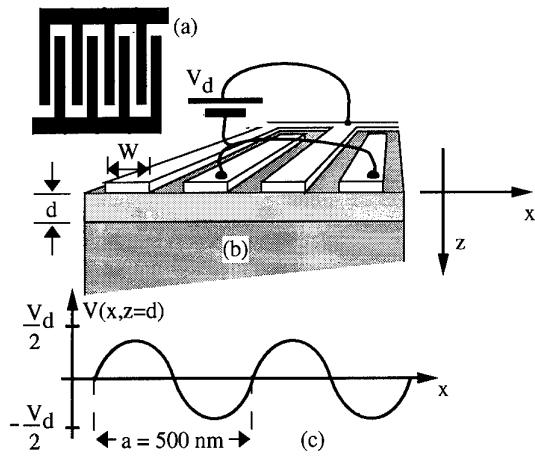


Figure 1. (a) Top view of an interdigitated gate. (b) Cross section of a heterojunction with an interdigitated Schottky gate. The voltage V_d is applied between the two electrodes of the gate so that at the location of the electron system a type II potential superlattice is formed as shown in (c). For instance, with $V_d = 1$ V the potential amplitude is 0.27 V at distance $d = 42$ nm below the surface.

defined with electron beam lithography on a bilayer PMMA resist. Each fingergate consists of 300 Ti stripes each 180 μm long and $W = 100$ nm wide. A superlattice potential of period $a = 500$ nm is induced if the finger-gates are biased with a voltage V_d with respect to each other. The amplitude of the potential superlattice decays exponentially into the z -direction normal to the crystal surface. However, if the distance d from the surface to the quantum well or the single heterojunction interface is small enough ($d \ll a$) a considerable portion of the modulation amplitude is still present at the electron system under investigation. Since in our metal-insulator-semiconductor (MISFET) type devices we do not use modulation doping in the front barrier region, there is no need for a thick spacer layer to preserve the high mobility of the electron system. Thus we can keep the distance from the gate to the electron system relatively low.

3. Capacitance of a quantum wire array in the one-dimensional quantum limit

In figure 2 a typical layer sequence is shown that is used for capacitance spectroscopy on one-dimensional wire arrays. The heterojunction contains a back electrode that is set back from the front barrier by a spacer layer of 100 nm undoped GaAs. As sketched in figure 2(b) the potential at the interface between the GaAs spacer and the front barrier can be controlled with the gate voltage V_g applied between the front electrodes and the

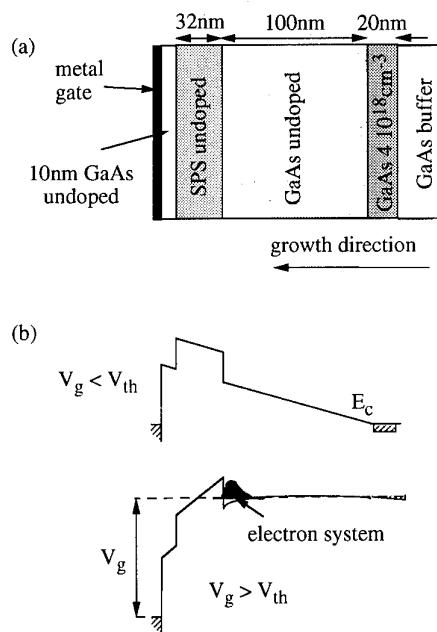


Figure 2. Layer sequence (a) of an MIS heterostructure with a back contact. The latter is formed by 20 nm Si^{n+} -doped GaAs grown on a GaAs buffer layer. The front barrier is formed by an AlAs/GaAs short-period superlattice (SPS) consisting of nine monolayers of AlAs and nine monolayers of GaAs respectively. (b) Conduction band onset as function of the z direction at gate voltage below and above the threshold voltage V_{th} respectively.

back contact. At sufficiently positive gate voltages $V_g \geq V_{th}$ the potential at the interface drops below the Fermi energy of the back electrode. The GaAs spacer forms a shallow tunnel barrier through which charge transfer from the back contact to the interface is possible. Thus a low-dimensional electron system is formed beneath the front gate electrode biased at $V_g \geq V_{th}$. The lateral geometry of the electron system is a replica of the geometry of this electrode. In samples with an interdigitated gate, one-dimensional electron wires are formed beneath one of the fingergates and the second fingergate, biased at $V_g - V_d$ with respect to the back electrode, controls the barriers between the wires.

The generation of the low-dimensional electron system at the heterojunction interface can be nicely monitored with the capacitance signal measured between the back contact and the front electrode. Here we measure the capacitance by the AC current through the sample excited with a small modulation of the gate voltage V_g of amplitude $dV \approx 5$ mV and frequency f . Whereas at $V_g < V_{th}$ the capacitance is determined by the geometry of the back and front electrodes alone, at $V_g \geq V_{th}$ the low-dimensional electron system contributes to the signal. It represents an additional electrode charged from the back contact through a resistance determined by the GaAs tunnel barrier. In samples with tunnel barriers of 100 nm thickness the charge exchange between back contact and the electron system takes place rapidly so that equilibration is obtained within microseconds. Thus even at $f = 100$ kHz the resistance of the tunnel barrier hardly affects the capacitance signal.

In figure 3 we present the capacitance of an electron wire array measured as function of the voltage V_g . Here the voltage difference between the fingergates is kept fixed at $V_d = 1$ V. At $V_g \leq 1.1$ V the signal reflects the capacitance between the back electrode and the front gate alone. At $V_g > 1.1$ V the capacitance rises rapidly, indicating that now one-dimensional wires are generated forming an electrode much closer to the front gate than the back contact. Numerical simulation calculations indicate that even at very small widths of the electron wires the capacitance measured at $V_g \geq V_{th}$ is predominantly determined by the capacitance between the front gate and the electron wires. It is clearly observable in figure 3 that this capacitance is not constant but exhibits pronounced steps as function of the gate voltage V_g .

These steps arise from the subsequent occupation of one-dimensional subbands in the electron wires. Since the electron wire array has a finite density of states (DOS) its capacitance is reduced below the value C_{geo} that a metallic electrode with infinite DOS would have. Because of the finite DOS of the wire array the charge increment in the wire array is accompanied by additional charges induced in the back electrode. The latter ensure that the electrochemical potential of the wire array stays constant although its Fermi energy increases. This fact can be approximately described by a capacitance $e^2 D(E)$ put in series with the geometrical capacitance: $C = (1/C_{geo} + 1/e^2 D(E))^{-1}$. The assignment of the steps to onsets of one-dimensional subbands is further verified

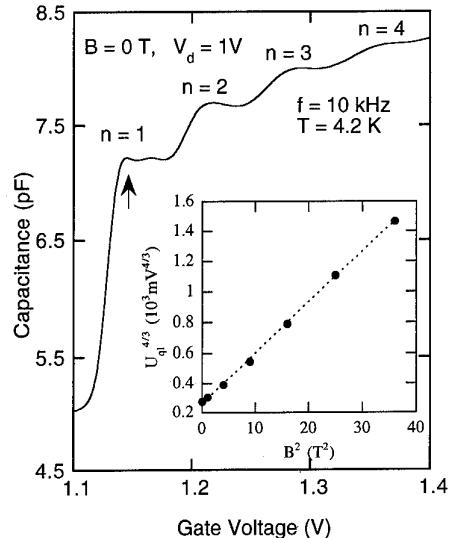


Figure 3. Capacitance of a wire array as function of the gate voltage V_g . The voltage between the interdigitated fingergates is kept fixed at $V_d = 1$ V. At the steps in the capacitance a new one-dimensional subband with index n starts to be occupied. The inset denotes the gate voltage position U_{q1} of the $n = 2$ step as function of a magnetic field applied perpendicularly to the wire array. The exponents are chosen such that equation (3) produces straight lines. The dotted line is a linear fit to the data with parameters given in the text.

by the dependence of the steps on the voltage V_d or on a magnetic field B applied perpendicularly to the wire array. With decreasing V_d the steps become less pronounced and their separation decreases as expected from a weaker confinement potential. In a magnetic field the gate voltages of the steps are clearly shifted to higher gate voltages, whereas the threshold voltage hardly changes (Drexler *et al* 1994a). This is expected since the density of states in one-dimensional subbands increases in a magnetic field.

Similar observations have been made on one-dimensional electron channels in shallow etched modulation-doped heterojunctions (Smith *et al* 1987, 1988a, b). There, however, structure in the capacitance signal arising from the subband occupation has been resolved only in the gate voltage derivative of the capacitance. In contrast, in our samples the steps are so strong that they can be easily resolved in the direct capacitance signal. This verifies the high homogeneity of our wire array. Furthermore, the capacitance signal rises very sharply at the onset of the first one-dimensional subband, indicating that even in the one-dimensional quantum limit a well defined quantum wire exists (Drexler *et al* 1994a). This, of course, is of particular interest because of the intriguing properties predicted for electron wires in the quantum limit (Kinaret and Lee 1990, Fasol and Sakaki 1992, Hu and DasSarma 1993, Schulz 1993, Shepard 1993).

The fact that here the direct capacitance signal is recorded makes comparison with self-consistent numerical calculations of the wire capacitance highly valid. However, such calculations still need to be performed. Here

we apply a simple model to derive parameters for the subband spacing and wire width. In this model we assume that the effective confinement potential as well as the wire capacitance C_w are constant at densities at which only the lowest subband is occupied. Correspondingly, the values thus derived for the subband spacing and wire width reflect averages for the density range at which the electron system is in the one-dimensional quantum limit. Furthermore, for the sake of simplicity, we assume that the effective confinement potential is of parabolic shape. At sufficiently low temperatures we have with these assumptions for the maximum one-dimensional electron density n_{q1} in the lowest subband at the onset of the occupation of the second subband

$$n_{q1} = \frac{C_w}{e} U_{q1} = \int_0^{E_{q1}} g_s \sqrt{\frac{2m^*}{\hbar^2} \frac{\sqrt{\omega_c^2 + \Omega^2}}{\Omega}} \frac{1}{\sqrt{E}} dE \quad (1)$$

where C_w is the wire capacitance, U_{q1} the voltage difference between threshold voltage and the voltage at the $n = 2$ capacitance step, $g_s = 2$ is the spin degeneracy factor, $\hbar\Omega$ the level spacing in the parabolic potential at $B = 0$, $E_{q1} = \hbar\sqrt{\omega_c^2 + \Omega^2}$ and $\omega_c = eB/m^*$ the cyclotron energy. From equation (1) we get a relation between the magnetic field B and the voltage difference U_{q1}

$$U_{q1}^{4/3} = a^{4/3}(\omega_c^2 + \Omega^2) \quad (2)$$

where $a = \sqrt{2m^*/\pi^2\hbar g_s/C_w\Omega}$. In the inset of figure 3 we depict the experimental values for U_{q1} versus the magnetic field together with a fit to the data according to equation (2). We find that the parameters $C_w = 140 \text{ pF m}^{-1}$ and $\hbar\Omega = 4.7 \text{ meV}$ best describe the experimental data.

For comparison we depict in figure 4 the capacitance calculated with $C = (1/C_{\text{geo}} + 1/e^2D(E))^{-1}$, with the density of states in a parabolic potential with 4.7 meV level spacing and assuming a constant C_{geo} , hence a constant wire width. The ideal dos has been convoluted with a Gaussian of 1 meV width to reflect broadening of the subband levels. Although the simple model applied here produces a capacitance of striking similarity we would like to point out two differences from the experimental data. The capacitance in figure 3 rises

almost monotonically whereas in the calculated capacitance we observe minima at the voltages at which the Fermi energy is in the regions of small dos. The voltage separations of subsequent capacitance steps in figure 3 stay roughly constant whereas the separations of the minima in the calculated spectrum increase with the subband index. Both observations are explained by the fact that in the experiments the wire width increases with increasing occupation. As expected the Coulomb repulsion of the electrons packed into the wire causes a renormalization of the effective confinement potential so that the wire becomes broader at higher electron density.

We observe a rapid renormalization of the effective confinement potential even at very low electron densities (Drexler *et al* 1994a). In figure 3 we resolve an additional structure in the capacitance signal highlighted with an arrow. The structure becomes even more pronounced in a magnetic field where it becomes obvious that the structure reflects an enhanced capacitance signal in a gate voltage region very close to the threshold voltage. We believe that this enhancement is caused by a rapid increase of the electron wire width in this gate voltage region. The corresponding increase of the wire capacitance results in an additional charge flow onto the capacitor and thus causes an enhanced capacitance signal (Drexler *et al* 1994a). This interpretation is augmented by measurements of the FIR absorption of the wire array. Here the capacitance signal is used to detect the dimensional resonance of the wire array (Drexler *et al* 1994b). At voltage $V_d = 1 \text{ V}$ we find the dimensional resonance at an energy of $\hbar\Omega_0 = 7.8 \text{ meV}$. According to the generalized Kohn theorem (Kohn 1961, Brey *et al* 1989) this energy corresponds to the single-particle level spacing in the bare confinement potential, i.e. the potential at the threshold voltage. This value is considerably higher than the previously derived average level spacing $\hbar\Omega = 4.7 \text{ meV}$. From twice the oscillator length we estimate a corresponding increase of the wire width from 24 nm to 30 nm, indicating a rapid increase of the wire width at electron densities in the one-dimensional quantum limit.

4. Lateral Franz-Keldysh effect in an InGaAs/GaAs quantum well

Important new aspects of the physics of potential superlattices in quantum wells are introduced in inter-band-optical experiments, because now electrons and holes are present in the system. Under these conditions the strong carrier pairing into excitons completely changes our view of the potential superlattice, because these quasiparticles sense electric field gradients rather than the fields themselves. Under sufficiently strong lateral fields, however, the excitons field-ionize into free electrons and holes. Nevertheless, the electron-hole correlation still exists in the sense that the single particles now occupy scattering states rather than bound exciton states. This is supported by our experiments on the lateral Franz-Keldysh effect, that are correctly described by

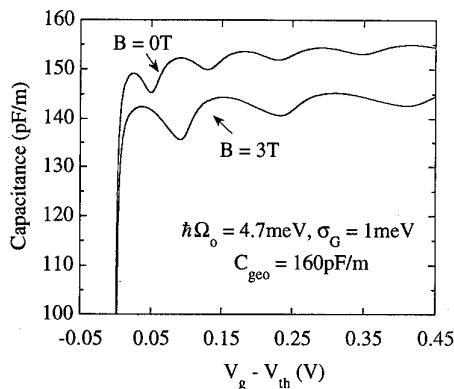


Figure 4. Capacitance of a wire array calculated according to a model described in the text. The trace for $B = 3 \text{ T}$ has an offset of -10 pF m^{-1} for clarity.

theories that take exciton effects into account (Schmeller *et al* 1994).

In order to investigate the optical properties of quantum wells in a pure type II potential superlattice we fabricated a molecular-beam-grown $\text{In}_x\text{Ga}_{1-x}\text{As}$ ($x = 0.15$, 10 nm thick) quantum well sandwiched in GaAs barriers. The GaAs barrier between the quantum well and the heterostructure surface was only 20 nm thick and the sample had no back electrode in order to minimize the field components in growth direction. As in the above experiments the metal stripes of the interdigitated gate had a width of 100 nm, and the gap between adjacent stripes was 150 nm. The samples were kept below 10 K and a tunable titanium-sapphire laser or monochromatized light from a halogen lamp was used for illumination. We calculate that the potential modulation in the plane of the quantum well is reduced only by $\approx 20\%$ with respect to the potential in the plane of the interdigitated electrode. We further find in the calculations that the lateral field component is about constant beneath the gaps between the stripes of the interdigitated gate. A small voltage applied between the finger gates suffices to induce lateral fields in these regions that field-ionize the quantum well excitons. We approximate the corresponding field from the ratio between the exciton binding energy R_y^* and the Bohr radius a_0^* of the exciton: $F_x \approx R_y^*/ea_0^*$. This field should be present in the gaps of the interdigitated gate at a voltage of about 0.2 V.

Indeed, we find that the luminescence intensity rapidly decreases in the gate voltage region $0 \leq V_d \leq 0.2$ V. At $V_d = 0.4$ V the luminescence intensity is about 20% of the intensity measured at $V_d = 0$ V and hardly changes at gate voltages beyond $V_d = 0.4$ V. Surprisingly, we find that the luminescence position (~ 1.37 eV) hardly shifts with the gate voltage. The decrease of the luminescence intensity indicates that the field-ionized electron-hole pairs no longer contribute to radiative recombination processes. However, it is not clear at present whether the luminescence remaining at voltages beyond $V_d = 0.4$ V is intrinsic to the potential superlattice or stems from unpatterned regions of the sample. Lineshape and position are identical to the luminescence from the unpatterned quantum well.

As demonstrated in figure 5, we measure the absorption of the quantum well by two different methods. The bottom spectrum in figure 5 is a conventional luminescence excitation spectrum (PLE), i.e. the luminescence intensity measured as function of the excitation energy. On the other hand, the top trace represents the photo-induced current through the interdigitated gate measured as a function of the excitation energy. Obviously, the current reflects the number of electron-hole pairs photogenerated in the quantum well. Nevertheless, there is an important difference between the two methods due to their different mechanisms. The PLE experiment measures the absorption in regions of the sample where the photogenerated carriers recombine radiatively with an energy at the detected position, whereas the photocurrent originates predominantly from electron-hole pairs ionized in the regions of strong lateral fields beneath

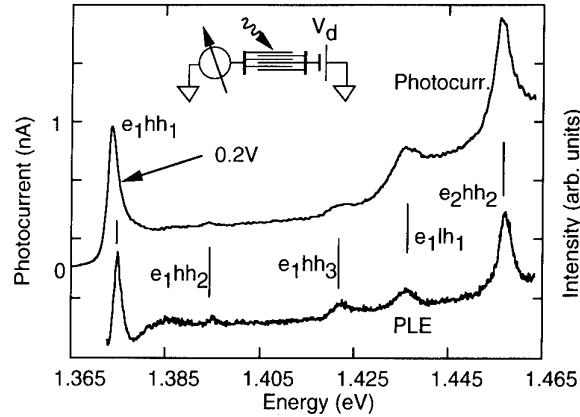


Figure 5. Photoluminescence excitation spectrum (PLE, right scale) and photocurrent spectrum (left scale) of an InGaAs/GaAs quantum well with interdigitated gate. The photocurrent trace is recorded with $V_d = 0.2$ V. The PLE trace is recorded on a part of the sample that is not covered by a gate. The notation identifying the excitonic features in the spectra refers to the two-dimensional subbands in the well. The inset sketches schematically the set-up for photocurrent measurements.

the gaps of the interdigitated gate. This explains why we find hardly any gate voltage dependence in the PLE spectra whereas in the photocurrent spectra we observe the effects of lateral fields on the quantum well absorption, namely the lateral Franz-Keldysh effect.

The main features are shown in figure 6–8. In figure 6 the photocurrent measured at larger voltages ($V_d \geq 0.2$ V)

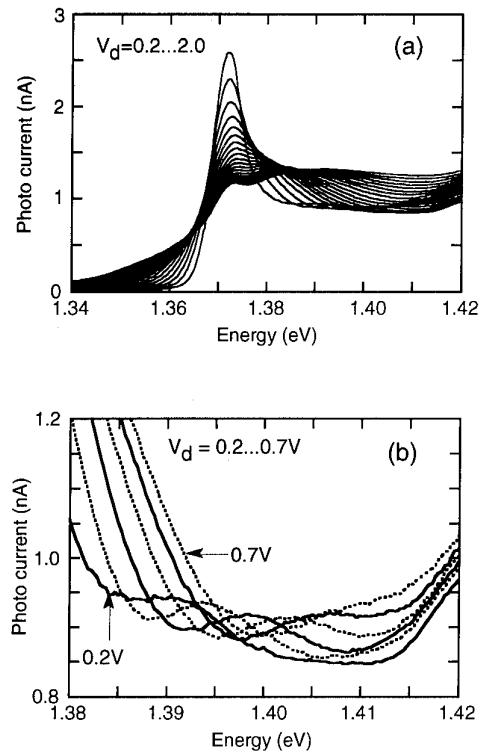


Figure 6. (a) Photocurrent spectra recorded on the sample of figure 5 at gate voltages $0.2 \leq V_d \leq 2.0$ V. (b) Photocurrent in a reduced energy range at moderate gate voltages between $V_d = 0.2$ V and 0.7 V.

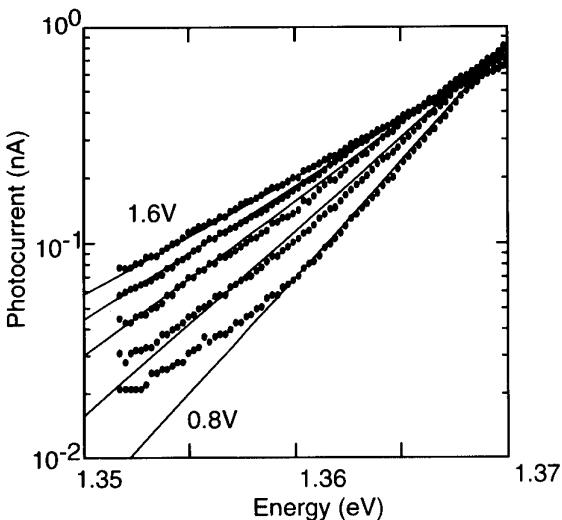


Figure 7. Logarithm of the photocurrent in figure 6 at gate voltages $0.8 \text{ V} \leq V_d \leq 1.6 \text{ V}$ versus the energy. The straight lines are calculated according to Dow and Redfield (1970) as described in the text.

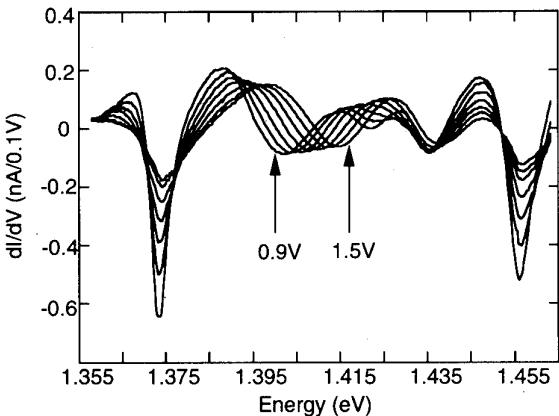


Figure 8. Gate voltage derivatives of photocurrent spectra at energies close to and above the effective bandgap of the InGaAs quantum well. Gate voltages between $0.9 \text{ V} \leq V_d \leq 1.5 \text{ V}$ in steps of 0.1 V have been applied to record the different traces. (From Schmeller *et al* (1994).)

is displayed as function of the energy close to the effective band gap energy E_g of the InGaAs quantum well. With increasing gate voltage we observe a growing low-energy absorption tail in the subgap region. Simultaneously, the excitonic feature at 1.373 eV that is very pronounced at small gate voltages diminishes and the absorption maximum shifts to higher energies. Furthermore, we observe an oscillatory structure (Tarmalingam 1963, Blossey and Handler 1972) in the energy dependence of the absorption at moderate gate voltages as shown in figure 6(b).

In figure 7 we plot the photocurrent measured at subgap energies in a logarithmic scale in order to compare the energy dependence with a theory by Dow and Redfield that includes the electron-hole correlation (Dow and Redfield 1970). The straight lines are calculated

according to $I(E) \sim \exp[c(E - E_g)/F]$ (Dow and Redfield 1970) where F is the lateral electric field and c a fit parameter common to all traces. Obviously, at high fields the absorption is very well explained with an exponential behaviour. Deviations at low fields might indicate a transition to non-exponential behaviour as is expected from theory.

Oscillations in the absorption as function of energy are best resolved in the gate voltage derivative of the absorption spectra (Schmeller *et al* 1994). This is demonstrated in figure 8, where the gate voltage derivative of the photocurrent is plotted as function of the excitation energy. The period of the oscillations increases with the lateral field. The absorption spectra show great similarities to theoretical predictions by Lederman and Dow for the lateral Franz-Keldysh effect in a homogeneous electric field applied parallel to an ideal two-dimensional system (Lederman and Dow 1976). So far a quantitative comparison with theory has not been attempted because it is numerically involved. In addition, we would like to point out that, in contrast to the model of Lederman and Dow (1976), in our superlattice the states involved in the absorption process are not free two-dimensional states but states in one-dimensional subbands or lateral minibands.

The fact that we probe the Franz-Keldysh effect in a lateral superlattice potential might also be the reason why the excitonic feature at 1.373 eV does not vanish completely in contrast to the predictions of Ledermann and Dow (1976). As in the case of the luminescence, the remaining structure hardly changes position and the decrease of its intensity saturates at high gate voltage differences.

Furthermore, we note that the oscillations show interference with the absorption into higher excitonic states at energies beyond 1.43 eV. Interference with the light-hole subband might have prevented the observation of Franz-Keldysh oscillations in previous experiments on GaAs quantum wells. In experiments by Miller *et al* (1985) on GaAs quantum wells a considerable lifetime broadening of the luminescence was observed but no oscillations in the absorption above bandgap.

The photocurrent excited at energies close to the bandgap energy shows a pronounced negative differential resistance very similar to the current-voltage characteristics of nipi structures (Döhler *et al* 1981) used for self-electro-optic-effect devices (SEEDS) (Miller *et al* 1984). We may thus expect similar potential for device applications for our structure. The small capacitance, the ease of fabrication as well as the orientation with respect to the plane of the quantum well could be advantageous in certain applications. The current-voltage characteristic depends on the illumination intensity, indicating that the lateral fields are partly screened by charge accumulation beneath the gate electrodes (Schmeller *et al* 1994). In samples with larger barriers between quantum well and front electrode, charge accumulation becomes even more important. Thus we may expect that an array consisting of alternating one-dimensional electron and hole plasmas is generated.

5. Conclusions

In this paper we present experimental results obtained on type II potential superlattices induced with very fine transducer gate electrodes. The amplitude of the superlattice potential is tuned with the gate voltage between the fingergates. The occupation of the electronic states in the superlattice can be controlled with a second gate voltage that is applied with respect to a back contact on the substrate side of a MISFET-type heterostructure. Charge injection into the superlattice takes place through a tunnel barrier separating the back electrode from the superlattice.

With this method we can, for example, induce very narrow one-dimensional electron wires in the minima of the potential superlattice. This is verified with capacitance measurements. We have precise control of the number of occupied one-dimensional subbands and can derive values of wire parameters such as subband spacing, electron density or wire width. The high homogeneity of the wire array even allows for investigations of wires in the one-dimensional quantum limit.

Interdigitated gates can equally well be employed for investigations of the optical properties of quantum wells in lateral type II superlattice potentials. With increasing amplitude of the potential superlattice the quantum well luminescence intensity diminishes considerably, reflecting the field-ionization of excitons. Whereas the absorption probed with luminescence excitation spectroscopy does not change significantly, the absorption probed with photocurrent spectroscopy changes drastically at high electric fields. The photocurrent technique predominantly probes the absorption in the high-field regions of the quantum well. The photocurrent spectra exhibit the signatures of the lateral Franz-Keldysh effect: a field-induced enhancement of the subgap absorption and oscillations of the absorption as function of energy. Accumulation of carriers beneath the gate electrodes may lead to the generation of photoinduced carrier plasmas of very high mobility even without the use of a back electrode.

Acknowledgment

We gratefully acknowledge financial funding of the Deutsche Forschungsgemeinschaft.

References

- Ashoori R C, Störmer H L, Weiner J S, Pfeiffer L N, Baldwin K W and West K W 1993 *Phys. Rev. Lett.* **71** 613
- Blossey D F and Handler P 1972 *Semiconductors and Semimetals* vol 9 ed R K Willardson and A C Beer (New York: Academic) p 257
- Brey L, Johnson N F and Halperin B I 1989 *Phys. Rev. B* **40** 10647
- Döhler G H, Künzel H, Olego D, Ploog K, Ruden P, Stoltz H J and Abstreiter G 1981 *Phys. Rev. Lett.* **47** 864
- Dow J D and Redfield D 1970 *Phys. Rev. B* **1** 3358
- Drexler H, Hansen W, Kotthaus J P, Holland M and Beaumont S P 1994b *Appl. Phys. Lett.* **64** 2270
- Drexler H, Hansen W, Manus S, Kotthaus J P, Holland M and Beaumont S P 1994a *Phys. Rev. B* **49** 14074
- Fasol G and Sakaki H 1992 *Solid State Commun.* **84** 77
- Franz W 1958 *Z. Naturf.* **13a** 484
- Hu B Y-K and DasSarma S 1993 *Phys. Rev. B* **48** 5469
- Kash K, Van der Gaag B P, Mahoney D D, Goetz A S, Florez L T, Harbison J P and Struge M D 1991 *Phys. Rev. Lett.* **67** 1326
- Keldysh L V 1958 *Sov. Phys.-JETP* **6** 763
- Kinaret J M and Lee P A 1990 *Phys. Rev. B* **42** 11768
- Kohn W 1961 *Phys. Rev.* **123** 1242
- Lederman F L and Dow J D 1976 *Phys. Rev. B* **13** 1633
- McEuen P L, Foxman E B, Meirav U, Kastner M A, Meir Y, Wingreen N S and Wind S J 1991 *Phys. Rev. Lett.* **66** 1926
- Miller D A B, Chemla D S, Damen T C, Gossard A C, Wiegmann W, Wood T H and Burrus C A 1984 *Appl. Phys. Lett.* **45** 13
- 1985 *Phys. Rev. B* **32** 1043
- Nixon J A and Davies J H 1990 *Phys. Rev. B* **41** 7929
- Schmeller A, Hansen W, Kotthaus J P, Tränkle G and Weimann G 1994 *Appl. Phys. Lett.* **64** 330
- Schulz H J 1993 *Phys. Rev. Lett.* **71** 1864
- Shepard K 1993 *Phys. Rev. B* **45** 13431
- Smith T P III, Arnot H, Hong J M, Knoedler C M, Laux S E and Schmid H 1987 *Phys. Rev. Lett.* **59** 2802
- Smith T P, Brum J A, Hong J M, Knoedler C M, Arnot H and Esaki L 1988b *Phys. Rev. Lett.* **88** 585
- Smith T P III, Lee K Y, Hong J M, Knoedler C M, Arnot H C and Kern D P 1988a *Phys. Rev. B* **38** 1558
- Tarmalingam K 1963 *Phys. Rev.* **130** 2204
- van Wees B J, van Houten H, Beenakker C W J, Williamson J G, Kouwenhoven L P, van der Marel D and Foxon C T 1988 *Phys. Rev. Lett.* **60** 848
- Wharam D A, Thornton T J, Newbury R, Pepper M, Ahmed H, Frost J E F, Hasko D G, Peacock D C, Ritchie D A and Jones G A C 1988 *J. Phys. C: Solid State Phys.* **21** L209

Optical study of II–VI semiconductor nanostructures

Le Si Dang[†], C Gourgon[†], N Magnea[‡], H Mariette[†] and C Vieu[§]

CEA–CNRS joint group 'Microstructures de Semiconducteurs II–VI'

[†] Laboratoire de Spectrométrie Physique—CNRS, Université Joseph

Fourier—Grenoble I, BP 87, F-38402 Saint Martin d'Hères Cedex, France

[‡] CEA/Département de Recherche Fondamentale sur la Matière Condensée,

SP2M/PSC, BP 85X, F-38041 Grenoble, France

[§] Laboratoire de Microstructure et de Microélectronique CNRS, 196 Avenue

H Ravera, F-92225 Bagneux, France

Abstract. In this paper we review some recent optical studies on CdTe-based nanostructures fabricated either by etching two-dimensional structures or by direct growth by molecular beam epitaxy. Photoluminescence can be observed on as-etched wires for sizes as small as 40 nm, due to the combined effect of exciton trapping on radiative centers and low surface damage. It is found that electrodynamic effects are very strong in gratings of wires for grating periods of about the photoluminescence wavelength.

1. Introduction

Current interest in semiconductor nanostructures is driven by the unique physical properties expected from the reduction of dimensionality in these systems but also by their potential for optoelectronic device applications [1]. So far much of the experimental research effort has been devoted to the fabrication of III–V semiconductor nanostructures, and only recently have reports appeared on ZnSe-, ZnTe- and HgTe-based nanostructures [2–4] (the latest exciting work on II–VI nanocrystallites can be found in [5, 6]). The most interesting result of these early studies is that surface damage produced by reactive ion etching in II–VIIs seem to be less severe than in III–Vs [2]. Thus it should be possible to carry out optical measurements on as-etched II–VI structures without the need for overgrowth of a protecting layer.

In this paper we review the optical properties of CdTe-based nanostructures fabricated either by dry etching two-dimensional (2D) quantum well (QW) structures or by direct epitaxial growth. For the as-etched nanostructures, the photoluminescence (PL) intensity at low temperature is found to be comparable to that of the initial 2D QW. However, the radiative decay is now dominated by the recombination of excitons localized at defect/impurities in the nanostructures. This localization mechanism is particularly efficient since it allows the observation of PL even in the smallest fabricated wires (40 nm). For the dots, surface damage is more severe so that the PL is completely quenched for sizes smaller than about 100 nm. Another interesting result is the observation of electrodynamic effects in gratings of wires. They take place for grating periods of the order of the emission wavelength (about 0.8

μm), inducing an increase in the PL intensity by a factor of 4 and in the polarization by a factor of 2 [7]. We also present preliminary optical results on two types of nanostructures fabricated directly by molecular beam epitaxy (MBE). The first type of nanostructure is ZnTe islands embedded in a CdTe matrix obtained by controlling the growth of ZnTe submonolayer on tilted substrate surfaces [8]. The second type is CdTe wires obtained by modulation of the in-plane lattice constant of (110)-CdTe/CdZnTe QWs grown over a (001)-CdTe/CdZnTe strained superlattice (SSL), following the method described in [9].

2. Experiment

The samples used for the fabrication of etched nanostructures are CdTe/ZnTe multiquantum wells (MQWs) grown by MBE on a (001)-GaAs substrate. The MQW structure contains five periods of CdTe (13 nm) and ZnTe (1.6 nm). It is grown on top of a thick and relaxed Cd_{0.84}Zn_{0.16}Te buffer layer, and capped by a 40 nm Cd_{0.84}Zn_{0.16}Te layer. Conventional electron beam lithography is used to pattern arrays of wires and dots on samples covered by a 150 nm thick PMMA resist. Patterns are transferred by Ar⁺ beam etching using a 50 nm Ti mask produced by a lift-off process. Each processed sample contains typically 200 regions (mesas) of 40 × 40 μm² corresponding to various lithography conditions (electron beam exposure time, spot size, etc). Under optimal fabrication conditions the smallest wire and dot sizes are 40 nm and 70 nm respectively (see

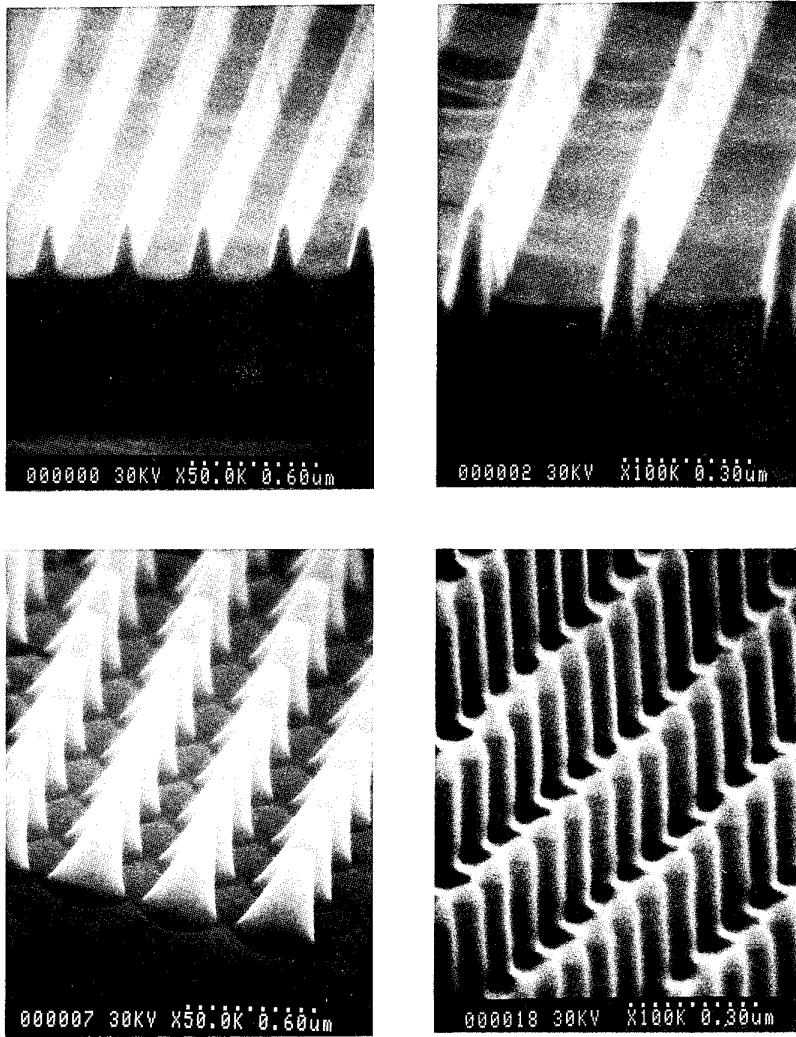


Figure 1. Scanning electron microscope images of periodic wires and dots etched into CdTe/ZnTe MQW. The MQW structure is capped by a layer of 100 nm thick CdZnTe. The smallest lateral sizes are about 40 nm for wires and 70 nm for dots.

figure 1). The etch depth is typically 300 nm, with sidewalls inclined by about 10°.

Nanostructures have also been grown directly by MBE using two different methods. In the first method, a superlattice of 0.5 monolayer (ML) of ZnTe and 30 MLs of CdTe is grown at 340 °C on a (001)-CdTe substrate tilted by 2° towards [100]. The choice of a rather high growth temperature is to ensure a growth through nucleation at step edges, thus favouring the formation of ZnTe islands of 1 ML height (see [8] for more growth details). The second method is a two-step epitaxial growth, as described in [9]. First, a (001)-CdTe/Cd_{0.77}Zn_{0.23}Te SSL is coherently grown on a Cd_{0.88}Zn_{0.12}Te substrate and capped by a 20 nm thick Cd_{0.77}Zn_{0.23}Te layer. The SSL contains 100 periods of CdTe (10 nm) and Cd_{0.77}Zn_{0.23}Te (10 nm). Then the sample is transferred into another chamber to be cleaved under a nitrogen atmosphere. This is to prepare the second growth step on the (110) cleaved facet by depositing a sequence of Cd_{0.77}Zn_{0.23}Te (20 nm), CdTe (5 nm) and Cd_{0.77}Zn_{0.23}Te (40 nm). Therefore

the bandgap of the (110)-CdTe/CdZnTe QW is spatially modulated by strains in regions grown above the (001)-CdTe/CdZnTe SSL, providing a lateral confinement to carriers to form quantum well wires (QWWs).

Optical measurements (PL, PL excitation, reflectivity) have been carried out at pumped liquid helium temperature (1.8 K). For PL measurements, a mapping set-up can be used to select any mesa on samples with a spatial resolution of 15 μm. PL excitation is provided by an Ar⁺ laser or a Ti:sapphire laser with typical power densities of 500 W cm⁻².

3. Optical results

3.1. Etched nanostructures

Figure 2 shows PL spectra of CdTe/ZnTe wires with lateral sizes varying from 170 nm down to 40 nm, together with PL spectra of 210 nm dots and of the 2D QW detected from a reference unetched mesa. The

reference 2D spectrum exhibits two sharp lines, X and Y. The X line is due to the intrinsic heavy-hole exciton $e_1 h_1$, and the Y line due to an exciton trapped on residual impurities or defects [10]. A few more lines are observed at lower energies (1.57–1.58 eV), possibly related to defect complexes [11]. It should be noted that this reference spectrum is identical to that observed on unprocessed samples, which clearly indicates that electron beam lithography and lift-off process do not produce any (visible) change in PL properties. However, after etching, large variations in PL are observed, as shown in figure 2. The Y line is now the dominant radiative recombination process, particularly for small nanostructures. The strong decrease of the X line is due to etching damage, which provides an additional non-radiative decay channel to photoexcited carriers not already trapped at Y centres. Recently, the influence of non-radiative recombination at sidewalls has been investigated on QWs fabricated by wet chemical etching of CdTe/CdMgTe single QWs [12]. In this process, etching damage is expected to be much less than in a dry etching process. However, PL results are no better: free and bound exciton recombinations are found to decrease for wire sizes smaller than 1 μm and about 250 nm, respectively. The latter value is comparable to the value of 100 nm observed for the Y line in figure 3(a). This result may seem surprising but it has been pointed out in [11] that, after implantation, optical properties are much poorer for single QWs than for MQWs. The reason is that implantation defects are trapped at the outer interfaces of the MQW, and thus the optical quality of the inner QWs is preserved. This defect trapping mechanism may also occur in our etched MQW samples. Another indication of etching damage in our samples can be seen in figure 2:

(i) both X and Y lines are blue-shifted. This shift cannot be due to the additional lateral confinement because it is observed for sizes as large as 170 nm for wires and 210 nm for dots;

(ii) the low energy defect lines at 1.57–1.58 eV are quenched.

All the above observations are similar to those reported in a previous annealing study of CdTe/ZnTe QWs after Ar^+ , Cd^+ or Zn^+ implantation at room temperature [11]. In particular, it is shown that defect lines at 1.57–1.58 eV can be easily quenched by a moderate annealing. Our results suggest that some local heating is actually produced during the dry etching process, and intermixing can then occur via structural defects generated by a straggling effect under the Ti mask.

No clear spectral blue shift due to the lateral confinement effect is observed in our samples. This is not surprising for two reasons: first, small confinement effects because of large carrier effective masses in wide bandgap II-VI semiconductors (about 0.1 for electrons and 0.6 for heavy holes in CdTe); second, broadening of optical transitions induced by inhomogeneous etching [13]. As shown in figure 2, PL (as well as PL excitation)

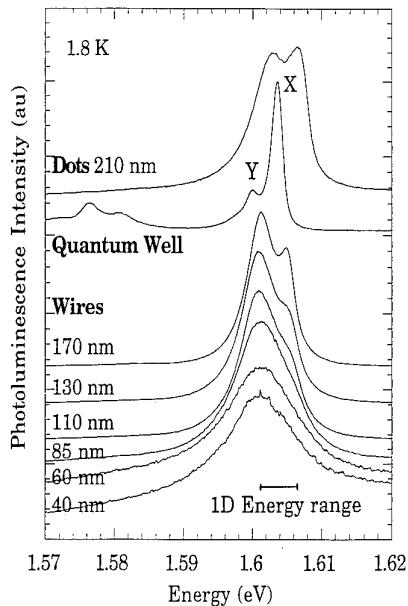


Figure 2. Photoluminescence spectra of arrays of wires with sizes varying from 170 nm down to 40 nm, together with spectra of arrays of 210 nm dots, and of a reference 2D qw. All spectra are normalized to their maximum intensities. Experimental conditions: $T = 1.8 \text{ K}$, non-resonant excitation at about 500 W cm^{-2} .

lines are broadened (up to 10 meV for the smallest wire size of 40 nm), and part of this broadening is due to the inclined sidewalls. For the MQW structure used in this work, an inclination of 10° would produce lateral wire sizes varying from 25 to 55 nm, and consequently an expected spectral blue shift varying from 2 to 7 meV (shown by the horizontal bar '1D' on figure 2). This will contribute to optical line broadening and make difficult a clear observation of lateral confinement effects.

Figure 3 shows a more quantitative analysis of the X and Y lines as a function of the lateral sizes of wires and dots. Their intensities have been determined by deconvolution of PL spectra obtained under resonant excitation into the CdTe QWs and subsequent normalization to the wire or dot surface. Note that PL results of 40 nm wires shown in figure 2 are not reported in figure 3 because they have been obtained on an array with a different period and under different excitation conditions. The normalized PL intensity is found to be weaker for dots than for wires of the same size. Moreover the X line is quenched for sizes smaller than 130 nm for dots, as compared with 100 nm for wires. These effects are due to etching damage, which is expected to be more severe for dots. In contrast to the case of GaAs/GaAlAs [13], no overgrowth is needed to observe PL from the smallest wires fabricated (40 nm). This is a combined effect of the use of a MQW structure, efficient trapping of photoexcited carriers by Y centres, and smaller losses from etch-induced surface states [2]. In figure 3, PL intensities of the X and Y lines of the reference 2D MQW are also shown for comparison. Their overall intensities are found to be greater than those of the biggest dots (210 nm) etched in this work. However, the reverse is true for wires

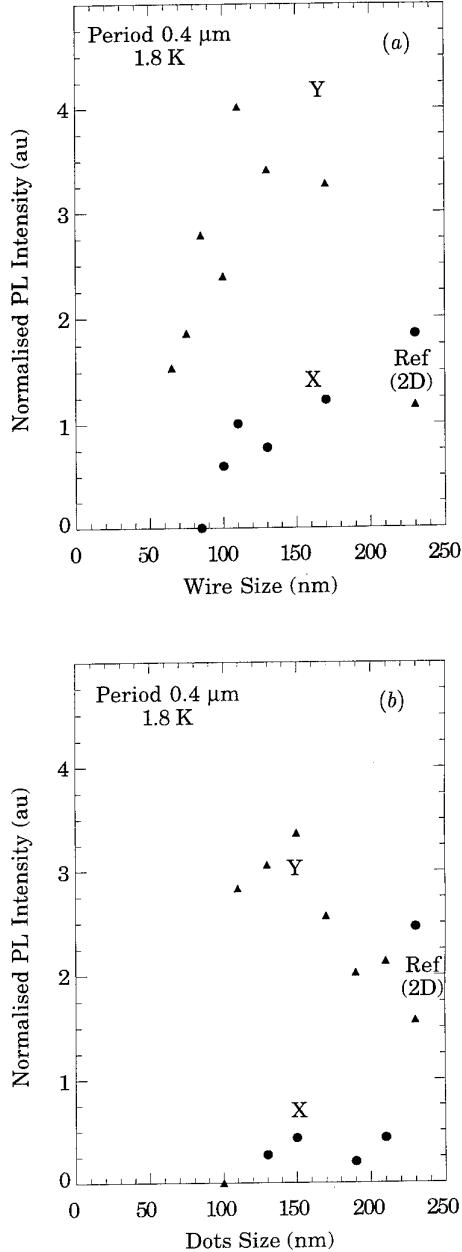


Figure 3. Photoluminescence intensities of free-exciton (X) and bound-exciton (Y) recombination as a function of the lateral sizes of wires and dots. The period of etched nanostructures is $0.4\text{ }\mu\text{m}$. Intensities are normalized to wire and dot surfaces. Results of the reference 2D qw are also shown for comparison. Experimental conditions: $T = 1.8\text{ K}$, resonant excitation into CdTe/ZnTe wires and dots at 500 W cm^{-2} .

(see figure 3(a)), which is surprising at first sight but can be explained by etch-induced annealing effects. As discussed above, quenching of the defect lines at $1.57\text{--}1.58\text{ eV}$ is strong evidence of annealing during the dry etching process. Therefore one would expect an increase of the PL efficiency by a factor of two for the X and Y lines [11], if etch damage can be neglected. In fact, the observed increase is only by about 50%, indicating that etch damage is quite significant, even for wires as wide as 170 nm .

Optical anisotropy is expected from quantum wires,

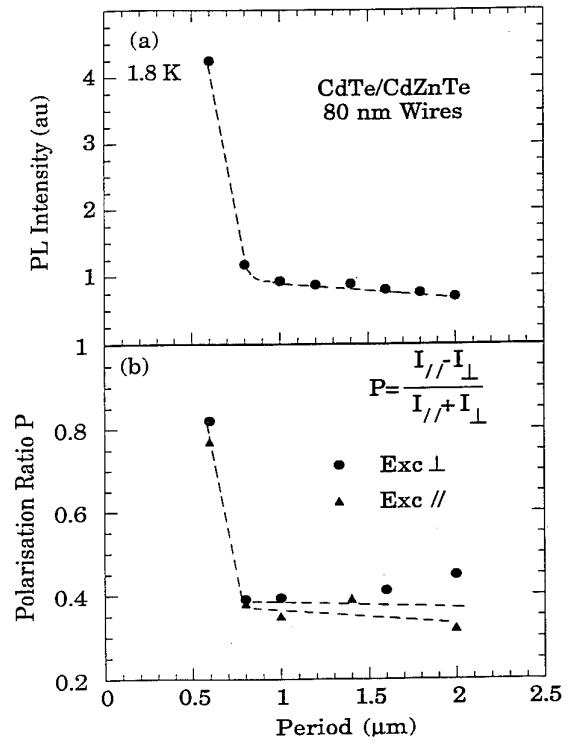


Figure 4. Photoluminescence of gratings of 80 nm wires as a function of grating period: (a) unpolarized intensity; (b) polarization ratio of the photoluminescence for excitation polarization parallel and perpendicular to the wire direction. Experimental conditions: $T = 1.8\text{ K}$, non-resonant excitation at 500 W cm^{-2} .

due to confinement-induced valence band mixing [14]. In the case of as-etched wires, the surface is corrugated and a strong polarization effect can also occur as a result of electrodynamic effects [15]. Figure 4 shows a study of PL intensity and polarization from gratings of 80 nm wires of different periods. There is a strong effect occurring for grating periods of about $0.6\text{--}0.8\text{ }\mu\text{m}$, which correspond to the emission wavelength ($0.77\text{ }\mu\text{m}$). PL polarization is strongly TE mode (up to 80%), independently of the excitation polarization. A confinement effect can be ruled out here since the wire size is kept constant throughout this study.

3.2. MBE-grown nanostructures

Figure 5 shows the reflectance spectra of two CdTe/ZnTe superlattices grown on (001)-CdTe substrates tilted by about 2° towards [100]. The thickness of CdTe is 30 MLs, and that of ZnTe is 0.5 ML in one sample and 2 MLs in another sample. The strain configuration in the superlattice is such that electrons and heavy holes are confined in CdTe and light holes in ZnTe. For the 2 MLs sample (figure 5(a)), we observe three lines which fall close to the expected positions of the heavy-hole excitons $e_1 h_1$ in CdTe/ZnTe superlattices with 1, 2 and 3 MLs of ZnTe. The same is true for the light-hole exciton $e_1 l_1$. The observation of peaks associated with integral numbers of ZnTe MLs indicates that interface smoothness extends on a scale larger than the exciton mean free path. For the 0.5 ML sample (figure 5(b)) the number of peaks

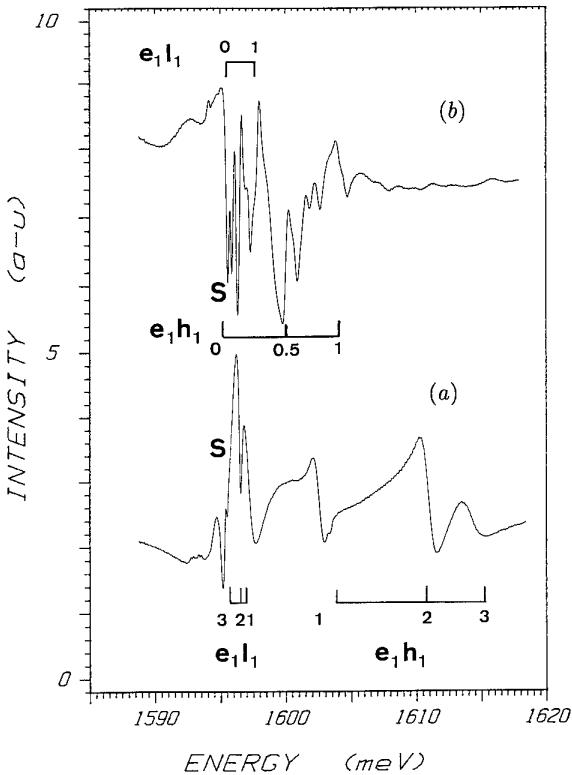


Figure 5. Reflectance spectra at 1.8 K of two (001)-CdTe/ZnTe superlattices grown by MBE on a tilted CdTe substrate. The thickness of CdTe is 30 monolayers, and that of ZnTe is 2 monolayers (spectrum (a)) and 0.5 monolayer (spectrum (b)) (see text).

in the reflectivity spectrum is even greater! They span the whole spectral range expected for heavy-hole and light-hole excitons in CdTe/ZnTe SSLs with 0 and 1 ML of ZnTe. Therefore it is tempting to attribute the various peaks to exciton localization induced by ZnTe islands of 1 ML height and different sizes, randomly distributed in the growth plane. Further optical studies (PL, PL excitation) are required to clarify this point.

Typical PL spectra are presented in figure 6 for the (110)-CdTe/CdZnTe QWWs fabricated by a two-step epitaxial growth. These PL spectra are obtained from three different regions of the sample, as indicated in the inset of figure 6. Figure 6(a) corresponds to PL of the (001)-SSL, figure 6(b) to PL of (110)-QWWs (grown over the SSL) and figure 6(c) to PL of (110)-QW (grown over the CdZnTe substrate). The PL of the SSL and of the (110)-QW exhibit two lines due to recombination of intrinsic heavy-hole excitons (e_1h_1) and of excitons bound to residual impurities (C and Y). The PL of QWWs is shifted by about 3 meV to the red relative to the (110)-QW, in agreement with estimates of carrier localization by strain modulation in this structure [16]. Note that this localization is quite efficient since the PL of QWWs dominates over that of SSL when exciting into wire regions (region (b) in figure 6). This is a very encouraging result for future optical spectroscopy.

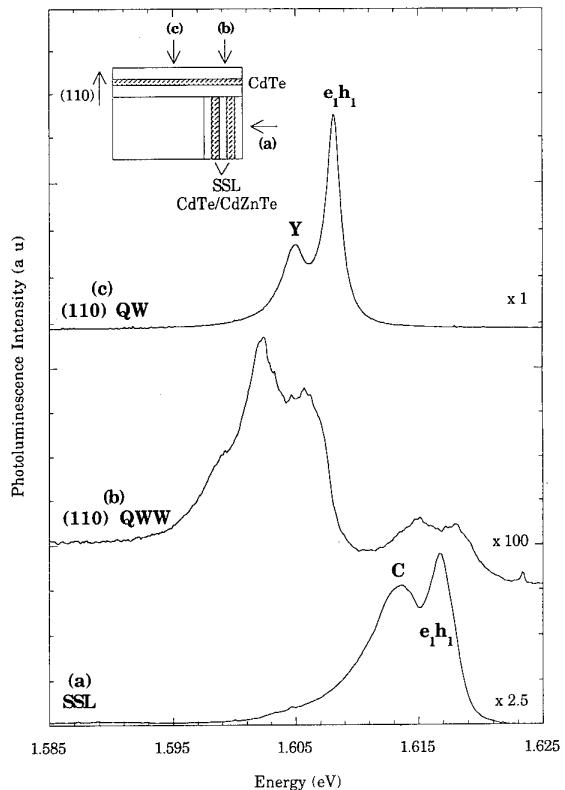


Figure 6. Photoluminescence spectra of the sample obtained by overgrowth of a (110)-CdTe/CdZnTe QW on a (001)-CdTe/CdZnTe strained superlattice. Spectra are obtained from three different regions on the sample, as indicated in the inset: (a) PL of (001)-superlattice; (b) PL of (110)-wires grown above the superlattice; (c) PL of (110)-QW grown above the substrate. Experimental conditions: $T = 1.8$ K, non-resonant excitation at 500 W cm^{-2} .

4. Conclusions

In summary, we have reviewed the optical properties of CdTe-based nanostructures fabricated either by dry etching 2D QWs or by direct MBE growth. It is found that no overgrowth is needed to observe PL from as-etched wires as small as 40 nm. This is a combined effect of exciton trapping on radiative centres and low losses from etching surface damage. The problem of surface damage is particularly severe for dots, and might be reduced by the use of reactive ion etching. We have also shown that strong polarization effects (up to 80% TE polarization) can be induced by electrodynamic effects in gratings of wires. We have presented preliminary optical results on nanostructures obtained by two methods: growth of a submonolayer on vicinal surfaces, and growth on a cleaved surface over a strained superlattice. The results are very promising considering the relative infancy of the field of II-VI nanostructures.

Acknowledgments

We wish to thank J Cibert, K Saminadayar and S Tatarenko for providing the MBE samples used in the dry etching work.

References

- [1] Weisbuch C and Vinter B 1991 *Quantum Semiconductor Structures* (San Diego: Academic)
- [2] Walecki W, Patterson W R, Nurmikko A V, Luo H, Samarth N, Furdyna J K, Kobayashi M, Durbin S and Gunshor R L 1990 *Appl. Phys. Lett.* **57** 2641
- [3] Foad M A *et al* 1991 *Proc. Int. Symp. on Nanostructures and Mesoscopic Systems, Santa Fe, 19–24 May 1991* (Amsterdam: North-Holland)
- [4] Eddy C R, Dobisz E A, Hoffman C A and Meyer J R 1993 *Appl. Phys. Lett.* **62** 2362
- [5] Bawendi M G 1993 *Proc. Sixth Int. Conf. on II-VI Compound Semiconductors, 13–17 September 1993, Newport, RI* (Amsterdam: North-Holland)
- [6] Bhargava R N, Gallagher D, Hong X and Nurmikko A 1994 *Phys. Rev. Lett.* **72** 416
- [7] Gourgon C, Eriksson B, Le Si Dang, Mariette H and Vieu C 1993 *Proc. Sixth Int. Conf. on II-VI Compounds Semiconductors, 13–17 September 1993, Newport, RI* (Amsterdam: North-Holland)
- [8] Magnea N 1993 *Proc. Sixth Int. Conf. on II-VI Compound Semiconductors, 13–17 September 1993, Newport, RI* (Amsterdam: North-Holland)
- [9] Gershoni D, Wiener J S, Chu S N G, Baraff G A, Vandenberg J M, Pfeiffer L N, West K, Logan R A and Tanbun-Ek T 1990 *Phys. Rev. Lett.* **65** 1631
- [10] Mariette H, Dal'bo F, Magnea N, Lentz G and Tuffigo H 1988 *Phys. Rev. B* **38** 12443
- [11] Hamoudi A, Cibert J, Feuillet G, Jouneau P H, Le Si Dang, Ligeon E, Pautrat J L, Saminadayar K and Tatarenko S 1993 *J. Appl. Phys.* **74** 2524
- [12] Illing M, Bacher G, Waag A, Forchel A and Landwehr G 1993 *Proc. Sixth Int. Conf. on II-VI Compounds Semiconductors, 13–17 September 1993, Newport, RI* (Amsterdam: North-Holland)
- [13] Birotheau L, Izrael A, Marzin J Y, Azoulay R, Thierry-Mieg V and Ladan F R 1992 *Appl. Phys. Lett.* **61** 3023
- [14] Bockelmann U and Bastard G 1991 *Europhys. Lett.* **15** 215
- [15] Bockelmann U 1991 *Europhys. Lett.* **16** 601
- [16] Fishman G 1994 private communications

Terahertz Bloch oscillations in semiconductor superlattices

T Dekorsy†, P Leisching†, C Waschke†, K Köhler‡, K Leo§,
H G Roskos† and H Kurz†

† Institut für Halbleitertechnik II, RWTH Aachen, D-52056 Aachen, Germany

‡ Fraunhofer-Institut für Angewandte Festkörperphysik, D-79108 Freiburg, Germany

§ Institut für angewandte Photophysik, Technische Universität Dresden, D-01062 Dresden, Germany

Abstract. The coherent excitation of excitonic wavepackets in AlGaAs/GaAs superlattices is investigated in a comparative study employing three different time-resolved optical techniques. The experimental results obtained in time-resolved four-wave mixing (FWM), THz emission spectroscopy and transmissive electro-optic sampling (TEOS) provide the final proof of the existence of Bloch oscillations in the THz range. Each of these measurements provides complementary information on these oscillations. As a result, Bloch oscillations appear as a special case of quantum interference involving the coherent superposition of Wannier–Stark states.

1. Introduction

Continuous progress in both molecular beam epitaxy and time-resolved optical techniques has established a firm base for directly studying quantum interference phenomena of excitonic wavepackets in semiconductor heterostructures in the time domain. Today, vertical nanostructures can be fabricated with atomic-layer precision. High-quality AlGaAs/GaAs heterostructures can be grown with an exciton linewidth of less than 1 meV at 10 K. As a consequence, the coherence of excitons is preserved for a couple of picoseconds, opening to the way to study quantum beat phenomena on a sub-picosecond time-scale. Superlattices have recently gained attention, because they allow one to study wavepackets in periodic potentials. At the centre of interest has been the postulation by Zener that electrons in a periodic potential subject to a uniform electric field should undergo time-periodic oscillations in real space [1], the so-called Bloch oscillations. Over two decades ago, Esaki and Tsu suggested that Bloch oscillations in semiconductor superlattices might serve as a source of frequency-tunable THz radiation [2]. The theoretical validity of the concept of Bloch oscillations and the experimental proof of their existence, however, remained a matter of great controversy for quite a time [3].

Recently, the simple quasi-classical picture of electrons moving in k -space in the electric field direction until they experience Bragg reflections at the Brillouin zone boundary has been replaced by a fully quantum-mechanical description [4]. Bloch oscillations appear as a special case of the quantum interference of Wannier–Stark states [5]. A realistic initial superposition

of these states predicts the time-periodic centre-of-mass oscillations as in the traditional semiclassical picture [3, 6].

Localization of Wannier–Stark states in AlGaAs/GaAs superlattices controls the degree of carrier localization [7]. Principally two different bias regimes can be distinguished. At low bias, the superlattice remains in the miniband regime; the wavefunctions of the electrons and holes are delocalized over a significant number of wells. The eigenstates form minibands. At large electric fields, each miniband splits into several equally spaced energy levels, forming a Wannier–Stark ladder. The energy separation of the Wannier–Stark states is larger than the scattering-controlled homogeneous linewidth. For increasing electric field, the wavefunctions become more and more localized until a complete two-dimensional behaviour is achieved. At moderate electric fields, the electron remains delocalized, while the heavy hole is strongly localized in a single well. The Coulomb interaction between delocalized electrons and localized holes plays a crucial role in the evolution of Bloch oscillations. Therefore, the direct measurement of time-periodic dipole moments associated with the displacement between electron and hole wavefunctions is of particular importance.

In transient time-resolved FWM experiments on biased GaAs/AlGaAs superlattices, periodic modulations in the decay of the diffracted signals have been assigned to Bloch oscillations [8–10]. The frequency of these oscillations follows the field/frequency relation predicted by both the semiclassical and the quantum-mechanical theory. In section 2, time-resolved FWM experiments are presented, which allow one to distinguish between quantum interference and polarization interference as

being the origin of the observed beating. However, FWM experiments cannot finally prove whether the oscillations are associated with a centre-of-mass motion of the electronic wavepacket. The symmetric excitation of Wannier-Stark states, e.g. of the heavy hole to -1 (hh_{-1}), hh_0 and hh_{+1} transitions with a broadband laser pulse, may be associated with a breathing mode, which does not result in an oscillating dipole moment.

A more direct proof of wavepackets oscillating in real space has been the observation of THz radiation emitted from the same superlattices after coherent excitation of several Wannier-Stark levels [11]. Phase, amplitude and frequency can be determined by optical gating of a subpicosecond photoconductive receiver antenna [11]. The measurements of time-dependent dipole radiation confirm the proposal of Esaki and Tsu [2]. Unlike FWM experiments, where interband polarizations are detected, the THz signal is related to the second time derivative of the intraband polarization $P_{\text{intra}}(\tau)$. Because of the coherent tunnelling of electrons back and forth across several superlattice periods, there will be a time-dependent dipole moment $P_{\text{intra}}(\tau) = Ne\langle z_e(\tau) \rangle$, where N is the number of dipoles and $z_e(\tau)$ the time-dependent displacement between optically generated electron and hole wavefunctions. In section 3 we briefly review the results obtained by this technique.

Further results on the temporal behaviour of the macroscopic polarization associated with Bloch oscillations are obtained in time-resolved electro-optic detection [12]. The phase and amplitude of the dipole can be detected up to more than 10 THz using appropriate polarization techniques. Transmissive electro-optic sampling (TEOS) can pick up periodic changes in the polarization from a relatively low background signal. Again, the intraband dynamics $P_{\text{intra}}(\tau)$ is observed to affect the polarization of a transmitted probe pulse through the electro-optic effect. Specific emphasis is paid to this latter technique in this article (section 4), since it appears to be powerful for the study of Bloch oscillations in the THz range.

In the following, the experimental results obtained with the three time-resolved techniques are compared for the case of the Wannier-Stark regime, i.e. the superlattices are externally biased until distinct Wannier-Stark transitions appear in absorption and photocurrent spectra. All investigations are focused on a GaAs/Al_{0.7}Ga_{0.3}As superlattice with 35 periods of 17 Å barriers and 97 Å wells. The lowest electronic miniband has a width of 19 meV. The heavy-hole wavefunction can be assumed to be completely localized in this regime. On top of the sample a semi-transparent Cr/Au Schottky contact is evaporated to enable the application of an electric field and the transmission of optical and THz beams. The superlattice is cooled to a lattice temperature of 10 K in all experiments. The laser source is a Kerr-Lens mode-locked Ti:sapphire laser emitting 150 fs (FWHM) transform-limited pulses. The excitation conditions differ slightly between the THz experiments and the all-optical experiments and they are noted specifically in the following sections.

2. Time-resolved four-wave mixing experiments

Up to now, most investigations of Bloch oscillations have been performed with time-integrated FWM experiments [8–10, 13]. In addition to the first discovery of Bloch oscillations, this method is specifically useful for studying their dependence on specific external parameters like electric field and temperature [13]. Measurement of the temporal evolution of the diffracted signal through nonlinear optical gating provides insight into many-particle interactions [14, 15]. This technique allows a clear distinction between homogeneous and inhomogeneous broadening to be made [16] and a clear answer as to whether Bloch oscillations arise from quantum interference or from polarization interference is possible [10, 17].

The temporal behaviour of the diffracted signal is described in a first approximation by a third-order density matrix approach for a non-interacting many-level model of the Wannier-Stark ladder [13]:

$$|\mathbf{P}_{\mathbf{k}_3}^{(3)}(t, \tau)|^2 = \Theta(t)\Theta(t - \tau) \exp[-2\gamma(t + \tau)] \times \left(\sum_k \mu_{0k}^8 + \sum_{k>j} 2\mu_{0j}^4 \mu_{0k}^4 \cos[\Delta\omega(k - j)t] \right) \quad (1)$$

where $\Theta(t)$ is the Heaviside function, τ the time delay between the two exciting laser pulses and t the time delay between the diffracted pulse and the probe pulse used for up-conversion. $\hbar\Delta\omega$ is the energy splitting between Wannier-Stark states. For simplicity we use $\gamma_{0k} = \gamma$ as the interband damping rate and assume a purely homogeneous broadening. The interband transition matrix elements are denoted by $\mu_{0k} = \langle 0_{hh} | k_{el} \rangle$. The most important results of equation (1) are the μ_{0k}^8 -dependence in the intensity of the diffracted light and the fact that the beating frequency of the FWM signal is given by $\Delta\omega$. Due to the μ_{0k}^8 -dependence, FWM is only sensitive to the heavy-hole transition of the Wannier-Stark ladder, which has a larger oscillator strength than the light-hole transition. The observation of a modulation of the FWM signal with $\Delta\omega$ clearly indicates the existence of a quantum beat phenomenon.

For the time-resolved experiments, the superlattice is biased electrically to the Wannier-Stark regime. A standard three-beam technique is employed: pump beam #2 is delayed by τ relative to the probe beam #1. The emitted FWM signal in the background-free direction of $\mathbf{k}_3 = 2\mathbf{k}_2 - \mathbf{k}_1$ is then up-converted with beam #3; the time delay between pulse #3 and #2 is denoted by t . The laser pulse is centred at 805 nm at the hh_{-1} Wannier-Stark transition. Under this condition, only the hh_{-1} and hh_0 transitions are excited by the broadband laser pulse at an electric field of 9 kV cm⁻¹.

In figure 1 the time-resolved FWM traces for different time delays τ between the two exciting pulses are shown for an excitation density of 3×10^9 cm⁻² carriers per superlattice period. The diffracted signal peaks after roughly 500 fs. The maximum remains

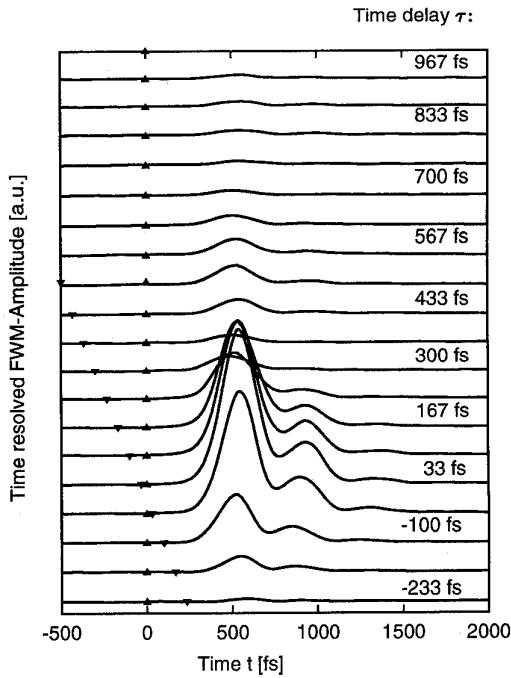


Figure 1. Time-resolved FWM signal for different delays τ at an excitation density of $3 \times 10^9 \text{ cm}^{-2}$ carriers. The triangles indicate the arrival of pulses #1 and #2. Pulse #2 arrives at $t = 0$ for all delays τ .

independent of the delay τ . No indication of a photon-echo-like behaviour is observed. From these results we conclude that the transitions involved are only homogeneously broadened. In both delay times t and τ , a beating with a frequency of about 400 fs is observed. This beating is unambiguous evidence that quantum interference of excited states is the mechanism responsible for the oscillations in the FWM signals. A polarization interference of independent Wannier–Stark levels would produce a different temporal behaviour [17]. The theory summarized in equation (1) does not include many-body Coulomb interactions and cannot explain the delayed signal, since in this simple theory the diffracted signal is emitted immediately after the second pulse (free polarization decay). The excitonic nature of Wannier–Stark transitions requires a much more elaborate theoretical treatment. The interaction of excitons has to be taken into account to interpret time-resolved FWM signatures correctly. It is known from previous time-resolved degenerate FWM experiments in single quantum wells that Coulomb-mediated many-body effects are able to explain the delayed rise of the diffracted signal [14, 15]. Polarization-wave interaction has a temporal dependence fundamentally different from that due to electromagnetic field interaction. Strong Coulomb interaction between excitons leads to the evolution of a polarization wave with a time delay of the order of the phase relaxation time T_2 , which is about 600 fs at this excitation density and temperature [13–15]. The rise time at negative delays τ is about half the decay time for the positive delays, as one expects from coherent polarization interaction in coupled quantum wells [18].

The quantum beats observed in FWM are denoted as Bloch oscillations for two reasons: (i) Differential

transmission experiments performed under the same excitation conditions as the FWM experiments show that the Wannier–Stark states with an energy separation that corresponds to the FWM beat frequency are excited. (ii) The bias dependence of the frequency is predicted by the semiclassical Bloch oscillation theory.

3. THz emission spectroscopy of Bloch oscillations

Time-resolved THz spectroscopy is a time-correlation measurement technique, which allows the detection of electromagnetic transients from photoexcited semiconductors [19–21]. Both phase and amplitude of coherent electromagnetic pulses are measured in the time domain up to THz frequencies. The typical detection bandwidth lies between 100 GHz and 2.5 THz. With an improved design of the receiver antenna, it is possible to extend the frequency range up to 4 THz. To perform the crucial experiments to decide whether Bloch oscillations emit dipole radiation, the same superlattice sample has been used as for FWM and TEOS measurements [11, 22].

The laser pulse width of about 150 fs corresponds to a spectral bandwidth of 18 meV. In the emission experiment, the wavelength of the laser is tuned to 800 nm, which is slightly higher in energy than for the case of FWM experiments. The laser beam is split into two beams. One beam excites the sample at an angle of 45° to the surface normal, while the other beam gates the receiver antenna. The excited far-infrared radiation is emitted through a semitransparent Schottky contact collinearly to the reflected optical laser beam. It travels several cm through air and is focused with paraboloidal mirrors onto a submillimetre photoconductive dipole antenna for detection. The antenna current is traced as a function of time delay between pump and gate pulses. The spot size of 350 μm is much larger than in the corresponding all-optical experiments. At an excitation density of approximately $5 \times 10^9 \text{ cm}^{-2}$ per well the excitation is slightly stronger than in the case of FWM and TEOS experiments. As a consequence, a significant offset of the reverse bias due to screening by accumulated charges has to be taken into account. A similar offset is observed for all experimental techniques.

In figure 2 the amplitude of the detected transient electric field is displayed as a function of time delay for a whole series of bias voltages. The separation between curves is proportional to the applied bias. The zero time delay is chosen arbitrarily. The emitted electric field transient contains contributions from both an instantaneous [11, 23] and a periodic intraband polarization. Above a bias of -1.2 V , the period of the oscillations starts to depend linearly on the applied voltage, and hence on the electric bias field, which is a clear indication of Bloch-oscillation-induced THz wave emission. Due to the limitation of the antenna bandwidth, the Bloch oscillation can only be followed up to 4 THz. The field dependence of the frequency is particularly visible in the Fourier transform of the

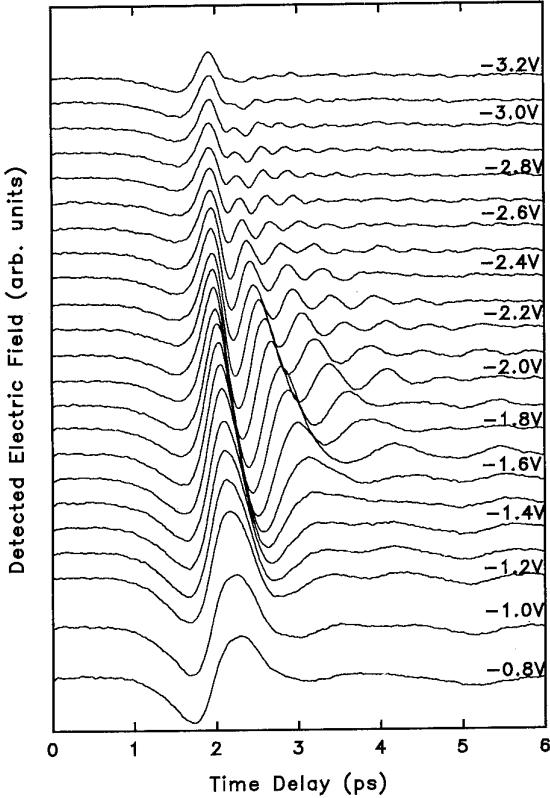


Figure 2. Terahertz transients from the semiconductor superlattice biased to the Wannier–Stark regime. The data are similar to those published in reference [18] for a different range of applied voltages.

time-resolved data [11, 22]. The frequency tuning can be traced over more than one order of magnitude from 0.2 THz to 4 THz [21]. The radiation can be observed up to 100 K at a miniband width of 19 meV [23]. As mentioned above, these results are a direct proof of the dynamical dipole behaviour of Bloch oscillations and greatly expand our understanding of the physics involved.

4. Electro-optic detection of Bloch oscillations

TEOS gives an insight into Bloch oscillation dynamics that is complementary to the techniques described above. The advantage of TEOS is the detection of the intraband polarization P_{intra} associated with Bloch oscillations with the time resolution of the optical pulse [12]. It has a temporal resolution superior to both FWM, which is restricted by the decay of the third-order interband polarization $P^{(3)}$, and THz experiments, which are restricted by the high-frequency cut-off of the antenna. In the THz experiment, the second time derivative of the polarization is detected in the far-field, while in the TEOS measurements the electric field is detected directly within the sample.

In the GaAs/AlGaAs superlattice of 43 m symmetry material the oscillating electric field induces birefringence, which is detected in a pump-probe set-up [24]. In the presence of an electric field, an anisotropic phase

retardation within the sample turns the transmitted probe beam of circular polarization into an elliptically polarized state. The degree of ellipticity is measured via a polarization analyser and two photodetectors. A similar technique has been applied for the observation of the macroscopic dipole moment associated with coherent LO phonons in GaAs [25]. The time delay is achieved with a retroreflector mounted onto a speaker driven at 100 Hz. This allows a real-time data acquisition with a sensitivity of about 10^{-7} relative transmission changes. For small electric fields, the TEOS signal is proportional to the change $\Delta E_z(\tau)$ of the electric field in the growth direction z of the superlattice:

$$\frac{\Delta T_{\text{eo}}(\tau)}{T_0} = \frac{4\pi}{\lambda} n_0^3 r_{41} \Delta E_z(\tau) d \quad (2)$$

where τ is the time delay, λ the wavelength, r_{41} the electro-optic coefficient, n_0 the isotropic refractive index and d the thickness of the superlattice region. r_{41} and n_0 have large dispersion at excitonic resonances in quantum well heterostructures. However, taking literature values for the bulk GaAs band edge, a sensitivity of 8.5 V cm^{-1} in ΔE_z can be estimated.

The electric field changes are related to the charge displacement associated with the coherent wavepacket oscillation via

$$\Delta E_z(\tau) = \frac{1}{\epsilon \epsilon_0} N e \langle \Delta z_e(\tau) \rangle = \frac{1}{\epsilon_0 \epsilon} P_{\text{intra}}(\tau) \quad (3)$$

where $e \langle \Delta z_e(\tau) \rangle$ is the microscopic dipole moment of the wavepacket and N the excitation density. $\langle \Delta z_e(\tau) \rangle$ is the displacement of the centre of mass of the electronic wavefunction relative to the corresponding heavy-hole centre of mass. The TEOS technique is closely related to the THz experiments, where the second time derivative of the polarization, $\partial^2 P_{\text{intra}} / \partial \tau^2$, determines the strength of the emitted electric field.

Figure 3 depicts the modulation of the anisotropic transmission of the probe pulse for different reverse bias

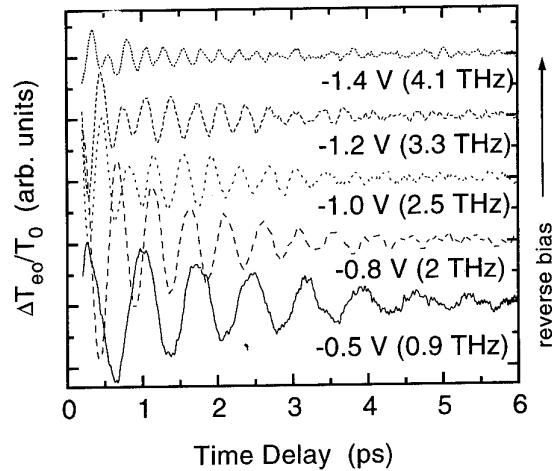


Figure 3. Oscillatory traces of anisotropic transmission changes of the superlattice for different applied reverse biases.

voltages. The sample is the same superlattice as used in FWM and THz emission with an electronic miniband width of 19 meV. The spectral position of the laser is centred at 805 nm. The excitation density in this case is $2 \times 10^9 \text{ cm}^{-2}$. At this low excitation level, the screening of the applied field due to accumulation of optically generated carriers is low, i.e. the formation of a Wannier–Stark ladder is already observed at -0.3 V reverse bias [13].

More than ten cycles of Bloch oscillations are clearly traced within the dephasing time of the intraband polarization. The signal resulting from the instantaneous polarization at zero time delay has been omitted. The increase of the Bloch oscillation frequency with increasing reverse bias is clearly visible. Another salient feature is the decrease of the initial Bloch oscillation amplitude with increasing applied field. It is directly related to a drop of the displacement amplitude $\Delta z_e(\tau)$ with reverse bias and is a clear indication of the increasing localization of Wannier–Stark states. We point out that the coherent wavepacket and the associated intraband polarization dephase within several picoseconds. This dephasing is significantly longer than the decay of the coherent third-order polarization detected in FWM experiments, which are additionally sensitive to the dephasing of holes.

One of the attractive features of Bloch oscillations in semiconductor superlattices is the wide range of tunability. At high frequencies, the tunability is restricted by the miniband width, while the low-frequency limit is given by the splitting of electronic states within the miniband. The frequency dependence of Bloch oscillations as determined from TEOS measurements is depicted in figure 4. The frequency can be tuned over more than one order of magnitude from 0.2 THz to 5.1 THz. This means that the complete range of frequencies that can be expected theoretically in a sample with 19 meV miniband width is covered in the high-sensitivity TEOS experiments. The lower frequency limit of 0.2 THz is below any energy splitting detectable in linear absorption spectroscopy.

In the Wannier–Stark regime a linear relation exists between frequency and applied electric field, as expected from the semiclassical Bloch oscillation theory. Slight deviations from this linear relation result from the change

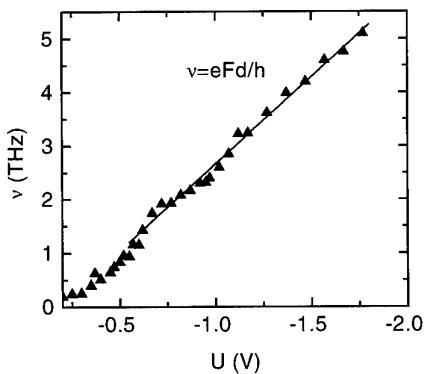


Figure 4. Bloch oscillation frequency versus reverse bias obtained from Fourier transformation of TEOS signals.

in absorption when the transitions are tuned electrically and the wavelength of the laser is kept constant. Thus, the density of optically generated carriers is varied too, affecting the dependence of the internal electric field on the externally applied voltage through screening effects. Below a critical offset voltage (-0.3 V) the oscillation frequency is affected only weakly by external fields.

5. Conclusion

In summary, we have studied the wavepacket dynamics in a superlattice structure biased into the Wannier–Stark regime by three different time-resolved optical techniques. Principally, degenerate FWM experiments yield information on the nonlinear interband polarization $P^{(3)}$. The dephasing of the nonlinear polarization is periodically modulated in these experiments by quantum beats of Wannier–Stark states, i.e. Bloch oscillations. However, the visibility of quantum beats is restricted by the decay of the third-order polarization.

The picture of semiclassical Bloch oscillations is confirmed by time-resolved THz spectroscopy. The excitation of dipole radiation is clearly connected to the centre-of-mass motion of wavepackets excluding breathing modes. The experimental approach to Bloch oscillations that is least limited in time resolution is the direct observation of the spatial oscillation of the intraband polarization detected in time-resolved internal electro-optic measurements. The possibility of detecting the amplitude and phase of the dipole moment between electrons and holes opens the way to a much deeper understanding of the coherent dynamics in superlattices.

Acknowledgments

We thank W Beck and Y Dhaibi for experimental help and fruitful discussions. This work was supported by the ‘Stiftung Volkswagenwerke’, the ‘Deutsche Forschungsgemeinschaft’ and the ‘Alfried-Krupp-Stiftung’.

References

- [1] Zener C 1934 *Proc. R. Soc. A* **145** 523
- [2] Esaki L and Tsu R 1970 *IBM J. Res. Dev.* **61** 61
- [3] Krieger J B and Iafrate G J 1986 *Phys. Rev. B* **33** 5494
- [4] Bastard G and Fereirra R 1989 *Spectroscopy of Semiconductor Microstructures*, NATO ASI Series (New York: Plenum) p 333
- [5] von Plessen G and Thomas P 1992 *Phys. Rev. B* **45** 9185
- [6] Bouchard A M and Luban M 1993 *Phys. Rev. B* **47** 6815
- [7] Mendez E E, Agulló-Rueda F and Hong J M 1988 *Phys. Rev. Lett.* **60** 2426
- [8] Feldmann J, Leo K, Shah J, Miller D A B, Cunningham J E, Schmitt-Rink S, Meier T, von Plessen G, Schulze A and Thomas P 1992 *Phys. Rev. B* **46** 7252
- [9] Leo K, Haring Bolivar P, Brüggemann F, Schwedler R and Köhler K 1992 *Solid State Commun.* **84** 943

- [10] Leisching P, Waschke C, Beck W, Haring Bolivar P, Roskos H, Leo K, Kurz H and Köhler K 1993 *Coherent Optical Interactions in Semiconductors, NATO ASI Series* (New York: Plenum)
- [11] Waschke C, Roskos H G, Schwedler R, Leo K, Kurz H and Köhler K 1993 *Phys. Rev. Lett.* **70** 3319
- [12] Dekorsy T, Leisching P, Beck W, Dhaibi Y, Leo K, Köhler K and Kurz H 1994 *Int. Quantum Electronics Conf., Anaheim* Technical Digest
- [13] Leisching P, Haring Bolivar P, Beck W, Dhaibi Y, Leo K, Köhler K and Kurz H 1994 *Phys. Rev. B* **50** at press
- [14] Dai-Sik Kim, Shah J, Damen T C, Schäfer W, Jahnke F, Schmitt-Rink S and Köhler K 1992 *Phys. Rev. Lett.* **69** 2725
- [15] Weiss S, Mycek M-A, Bigot J-Y, Schmitt-Rink S and Chemla D S 1992 *Phys. Rev. Lett.* **69** 2685
- [16] Yajima T and Taira Y 1979 *J. Opt. Soc. Japan* **47** 1620
- [17] Koch M, Feldmann J, von Plessen G, Göbel E O and Thomas P 1992 *Phys. Rev. Lett.* **69** 3631
- [18] Leo K, Wegener M, Shah J, Chemla D S, Göbel E O, Damen T C, Schmitt-Rink S and Schäfer W 1989 *Phys. Rev. Lett.* **65** 1340
- [19] Roskos H G, Nuss M C, Shah J, Leo K, Miller D A B, Fox A M, Schmitt-Rink S and Köhler K 1992 *Phys. Rev. Lett.* **68** 2216
- [20] Planken P M C, Nuss M C, Knox H, Miller D A B and Goosen K W 1992 *Appl. Phys. Lett.* **61** 2009
- [21] Roskos H G 1993 *Proc. Ultrafast Processes in Spectroscopy, Vilnius, Lithuania* to be published
- [22] Waschke C, Leisching P, Haring Bolivar P, Schwedler R, Brüggeman F, Roskos H G, Leo K, Kurz H and Köhler K 1994 *Solid-State Electron.* **37** 1321
- [23] Waschke C, Roskos H G, Leo K, Kurz H and Köhler K 1994 *Semicond. Sci. Technol.* **9** 416
- [24] Ralph S E, Capasso F and Malik R J 1989 *Phys. Rev. Lett.* **63** 2272
- [25] Cho G C, Kütt W and Kurz H 1990 *Phys. Rev. Lett.* **65** 764

Coherent dynamics of exciton wavepackets in semiconductor heterostructures

J Feldmann[†], M Koch[†], E O Göbel[†], F Jahnke[‡], T Meier[†],
W Schäfer[§], P Thomas[†], S W Koch[†], H Nickel^{||}, S Luttgendt[†] and
W Stolz[†]

[†] Philipps-Universität Marburg, Physics Department and Material Sciences Center, 35032 Marburg, Germany
[‡] Optical Sciences Center, University of Arizona, Tucson, AZ 85721, USA
[§] Forschungszentrum Jülich, 52425 Jülich, Germany
^{||} Deutsche Bundespost, Telekom, Fernmeldetechnisches Zentralamt, 64295 Darmstadt, Germany

Abstract. The coherent dynamics of exciton wavepackets in (GaIn)As/GaAs as well as (GaIn)As/Ga(PAs) multiple quantum well structures is studied by means of transient four-wave mixing (FWM) experiments. The wavepackets are generated by simultaneous excitation of several exciton transitions with laser pulses of about 100 fs duration. The time-integrated FWM signals exhibit a pronounced modulation superimposed on the overall decay which can be attributed to the quantum interference of the different eigenstates. In the time-resolved FWM signals this interference is not present, reflecting the interplay between many-body Coulomb effects and inhomogeneous broadening. This experimental technique is then employed to extract the exciton binding energies in pseudomorphic symmetrically strained (GaIn)As/Ga(PAs) with various In contents.

1. Introduction

A wavepacket represents a coherent superposition of electronic eigenstates. These wavepackets exhibit characteristic dynamics due to the interference of the respective phase factors as long as the coherence is maintained. Due to the extended nature of the intrinsic eigenstates in semiconductors the loss of coherence in general is extremely fast and occurs typically on a picosecond time-scale. Experimental investigations of coherent wavepacket dynamics thus require subpicosecond time resolution.

In this paper we report on experimental studies of the coherent dynamics of exciton wavepackets in semiconductor quantum well structures. The wavepackets are created by excitation of exciton transitions with short, that is spectrally broad, laser pulses of approximately 110 fs duration. The subsequent dynamics are monitored by measuring the nonlinear signal generated by four-wave mixing (FWM) as a function of time delay between two subsequent excitation pulses. In some cases the nonlinear signal is also resolved in real time, providing additional information on the nature of the electronic states involved.

As will be demonstrated, these experiments not only provide insight into the intrinsic dynamics of wavepackets and the scattering processes which destroy the

phase coherence, but, in addition, allow us to extract material parameters such as the exciton binding energy. This is of particular importance for quantum structures where the determination of these quantities by means of linear optical spectroscopy is often very difficult due to the unavoidable inhomogeneous broadening of the optical transitions.

In this paper we discuss wavepackets formed by the superposition of different transitions belonging to one single exciton, that is the wavepacket is formed by the coherent superposition of the lowest (1s) and the excited exciton states including parts of the exciton continuum. Other prominent examples investigated recently are, for example, wavepackets in double quantum well structures [1, 2], superposition of heavy- and light-hole excitons [3, 4], and the superposition of Wannier–Stark states in a semiconductor superlattice resulting in Bloch oscillations [5–7].

2. Experiment

We have used time-resolved FWM in our studies. The experimental set-up is schematically depicted in figure 1. A Kerr lens mode-locked Ti:sapphire laser generating about 110 fs pulses at a repetition rate of 76 MHz is employed. This laser is tunable in the spectral regime of

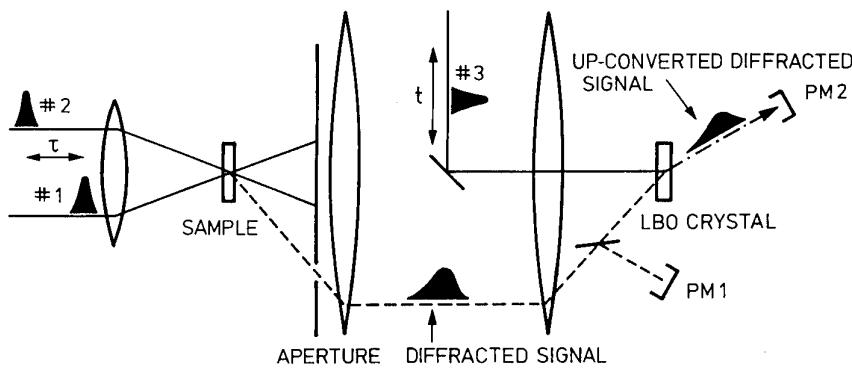


Figure 1. Experimental set-up for time-resolved and time-integrated FWM. For time-integrated detection the nonlinear signal is measured as a function of time delay between the excitation pulses by a photomultiplier (PM1). Time-resolved detection is achieved by up-conversion of the nonlinear signal in an LBO crystal.

0.7–1 μm . This tuning regime for the fundamental output can be considerably increased by means of nonlinear techniques like frequency doubling, parametric processes [8] or a combination of these. The laser output is split into two beams which are focused onto the sample, usually kept at liquid He temperatures, according to wavevectors k_1 and k_2 , respectively. These two pulse trains can be delayed with respect to each other by an optical delay line.

Due to nonlinear interaction, a signal is then generated in the direction $k_3 = 2k_2 - k_1$, provided the delay time τ is less than the dephasing time T_2 which is the characteristic time constant for the loss of coherence of the excited states. Note, however, that in many cases the coherence decay cannot be described by a constant rate corresponding to a simple time constant. In lowest order the nonlinear signal is created by the induced third-order polarization and is proportional to $|P^{(3)}(t, \tau)|^2$. This nonlinear signal can be detected either time integrated as a function of delay time τ by employing a ‘slow’ detector, for example, a photomultiplier (PM1), or time resolved for a series of fixed delay times τ by using an up-conversion scheme as also shown in figure 1. It should be mentioned that in addition to the intensity of the nonlinear signal, the phase of the emitted field with respect to the driving laser fields can be used to obtain information on the microscopic nature of the nonlinear interaction [9].

The proper theoretical description of these FWM experiments is based on the semiconductor Bloch equations [10, 11], representing a set of coupled differential equations describing the coupled dynamics of the population and polarization. These semiconductor Bloch equations include Coulomb interaction within the frame of the time-dependent Hartree–Fock approximation. It has been shown recently that in fact Coulomb interaction modifies the coherent dynamics considerably [12–15] as compared with atomic or molecular systems, which can be described by the ‘conventional’ optical Bloch equations. Nevertheless, in many cases the optical Bloch equations are applied to semiconductors because of their much simpler mathematical structure. The relevant

electronic transitions are then described within a multi-level scheme representing the excitonic resonances [16]. Even though this very often provides a qualitative understanding of the basic features observed in FWM experiments great care must be taken when quantitatively analysing experimental results in semiconductors.

3. Results and discussion

3.1. Exciton wavepackets in InGaAs/GaAs: effects of disorder and Coulomb interaction

We have recently reported [17] the observation of the coherent dynamics of exciton wavepackets in (GaIn)As/GaAs multi quantum well structures. As mentioned before, in these experiments all the bound as well as some of the continuum states of a quasi two-dimensional exciton are excited coherently by a spectrally broad laser pulse with spectral width exceeding the exciton binding energy. In the time domain this can be viewed as excitation of the exciton by a pulse with time duration much shorter than the semiclassical ‘orbit time’ of the electron around the hole. This situation for the excitation condition is illustrated schematically in the right-hand part of figure 2. The left-hand part shows the photoluminescence excitation (PLE) spectrum and the photoluminescence (PL) spectrum of the sample used in the experiment. The sample contains 20 $\text{In}_{0.08}\text{Ga}_{0.92}\text{As}$ quantum wells of 10 nm width separated by 40 nm wide GaAs layers. The PLE spectrum shows the lowest heavy-hole ($n = 1, 1s, hh$) as well as the lowest light-hole exciton ($n = 1, lh$) resonance. The linewidth (full width at half-maximum) of the heavy-hole exciton transition amounts to 1.7 meV. The Stokes shift between the heavy-hole emission line in PL and the corresponding absorption line in PLE is less than 0.2 meV, altogether demonstrating the high quality of the structure. The PLE spectrum, furthermore, shows a weaker peak corresponding to the 2s heavy-hole transition at 1.4611 eV. From its separation from the 1s peak, the binding energy of the exciton is determined to be approximately 7.5 meV.

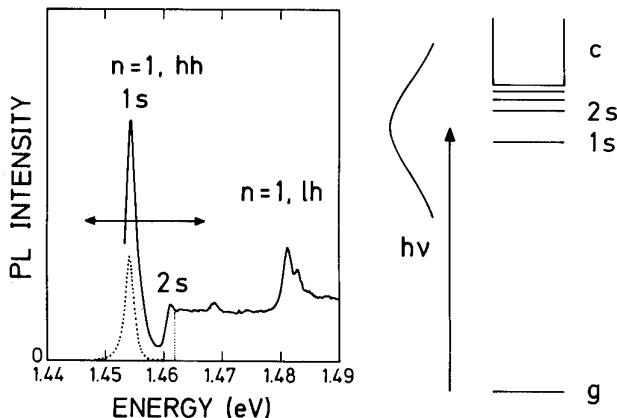


Figure 2. Photoluminescence excitation spectrum (full curve) and photoluminescence spectrum (dotted curve) of an $\text{In}_{0.08}\text{Ga}_{0.92}\text{As}/\text{GaAs}$ multi quantum well sample at $T = 5\text{ K}$ (left-hand side) and schematic representation of the excitation process using pulses of about 110 fs duration with the corresponding spectrum.

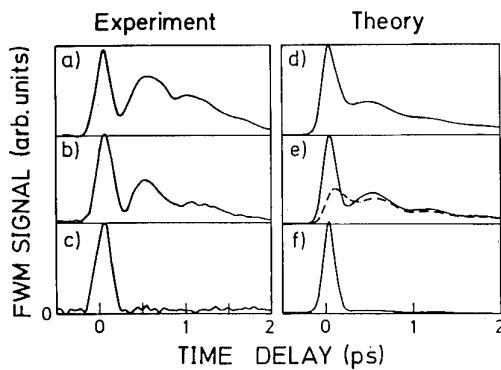


Figure 3. Low-temperature ($T = 5\text{ K}$) FWM signal as a function of delay time between the two excitation pulses for the $\text{In}_{0.08}\text{Ga}_{0.92}\text{As}/\text{GaAs}$ multi quantum well sample (cf figure 2). The experimental results are shown on the left and compared with calculations as described in the text (right-hand side). The different curves correspond to different detunings, Δ , of the central photon energy of the laser with respect to the 1s exciton resonance ($\Delta = -5\text{ meV}$ for curves (a) and (d), $\Delta = 0\text{ meV}$ for (b) and (e), and $\Delta = +10\text{ meV}$ for (c) and (f)).

The respective onset of the exciton continuum is indicated by the vertical dotted line. The large blue shift of the light-hole exciton is caused by the inherent strain in the GaInAs layers of this pseudomorphic structure. For our experiments this is of great advantage, since it allows us to excite solely the heavy-hole exciton states.

The normalized time-integrated FWM signal as a function of delay time τ is depicted in figures 3(a)–(c) for three different spectral positions of the excitation laser pulses. The central photon energy of the laser pulse spectrum is chosen to be exactly at the $n = 1$, 1s, hh exciton resonance (3(b)), 5 meV below (3(a)), and 10 meV above (3(c)) the resonance. The exciton density is kept low and amounts to about $4 \times 10^9 \text{ cm}^{-2}$. In either case, a fast increase of the FWM signal followed by a fast initial decrease is observed. For excitation below as well as at

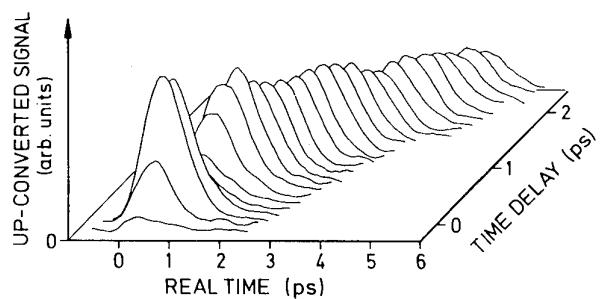


Figure 4. Time-resolved FWM signal for different time delays for the sample shown in figure 3.

the exciton resonance (figures 3(a) and (b)) the signal recovers again and reaches a second peak at $\tau = 550\text{ fs}$. The subsequent decay of the FWM signal is weakly modulated with a period of $T_b = 600\text{ fs}$, which actually corresponds to an energetic separation of 6.9 meV. This value coincides precisely with the 1s–2s separation as determined from PLE. It is thus concluded, and further supported by the calculations discussed below, that this modulation reflects quantum beats of the 1s and 2s exciton states. On the other hand, the initial fast decay of the signal is caused by an interference of the 1s exciton transition with its continuum transitions.

The experiments have been simulated using the phenomenological multilevel optical Bloch equations including finite pulse width. The basis for these calculations is the energy level scheme depicted on the right-hand side of figure 2. The energetic positions of the 1s and 2s excitons and the continuum transitions, as well as their respective optical dipole matrix elements, are chosen to fit the PLE spectrum. The continuum is modelled by closely spaced discrete levels with spacing of 0.5 meV. The dephasing times for the 1s and 2s excitons are taken to be 2 ps and 1 ps respectively, while for the continuum states a dephasing time constant of 400 fs is assumed. The dependence on delay time of the FWM signal is calculated for the same spectral positions of the excitation laser as in the experiment and is shown on the right-hand side of figure 3(d)–(f). The calculated curves reproduce the time-integrated experimental result quite well. In particular, the initial fast decay can be identified to be due to an interference effect between the 1s exciton transition and the continuum transitions. This can be seen by the broken curve in figure 3(e), where the continuum states have been neglected. Clearly this curve reproduces the 1s–2s beating as expected; however, it does not provide a proper description of the initial dynamics.

Since the phenomenological optical Bloch equations neglecting inhomogeneous broadening as well as Coulomb interaction seem to describe the results of the FWM experiment for time-integrated detection fairly well, one would expect the same qualitative agreement for time-resolved detection, that is for the time behaviour of $|P^{(3)}(k_3, t; \tau)|^2$. However, this is not the case, demonstrating once again the importance of many-particle Coulomb interaction in the coherent dynamics. In figure 4 we show

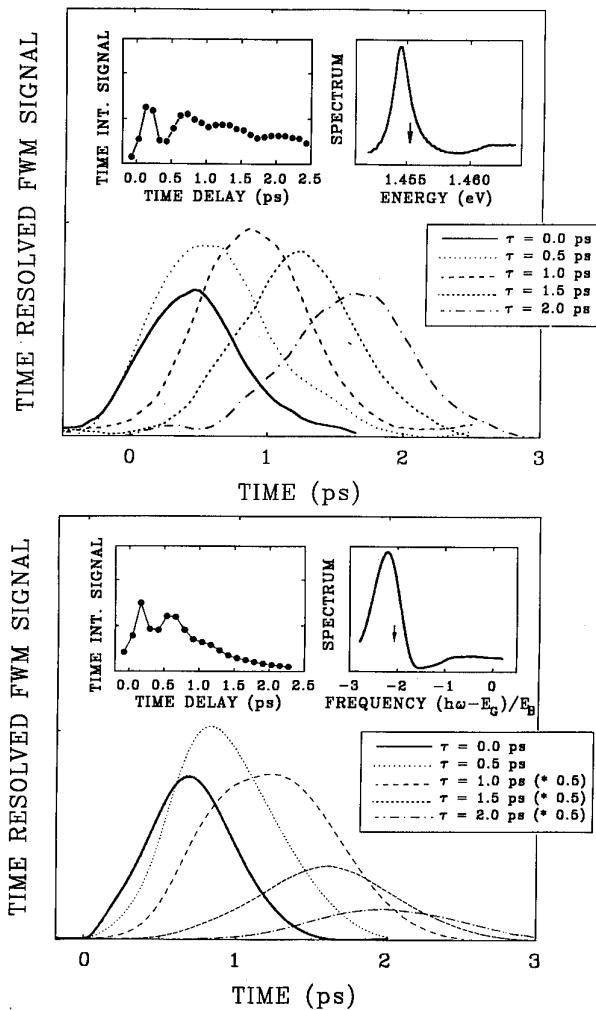


Figure 5. Comparison of experimentally determined time-resolved FWM signals at different time delays (upper part) with calculated curves employing the semiconductor Bloch equations (lower part). Also shown are the time-integrated FWM signals (upper left insets) and the linear absorption spectrum (upper right insets).

the FWM signal as a function of real-time t for different delay times τ . For a fixed value of τ , the FWM signal as a function of real-time does not exhibit a beating behaviour, in contrast to the predictions of the simple Bloch equations used above. The maxima of the individual FWM curves are located approximately along $t = 2\tau$ as characteristic for photon echoes, that is for strongly inhomogeneously broadened transitions. However, one has to bear in mind that Coulomb interaction also leads to a delayed real-time increase of the FWM signal [13–15]. It also has to be noted that the temporal width of the echo-like FWM signal increases with delay time τ which again cannot be explained within the multilevel optical Bloch equations. We therefore have extended the theoretical analysis by applying the full semiconductor Bloch equations. The results are shown in figure 5 (lower part) and compared with the experimental results depicted once again in the upper part of figure 5. The insets in the upper left and right respectively, show the

time-integrated FWM signal as well as the linear absorption spectrum. The centre frequency of the excitation laser is indicated in the absorption spectrum by a vertical arrow. The main parts of the figures depict the time-resolved FWM signal for different delay times. The overall agreement between experiment and theory is reasonably good. In particular, the calculated time-resolved FWM traces do not show a modulation and correspond to the measured curves. An analysis of the calculated curves shows that the real-time signal consists of a prompt and a delayed interaction-induced part which basically reflect, respectively, the electric field term and the polarization term of the renormalized Rabi frequency [13]. The numerical calculations further show that a quantitative reproduction of the experimentally observed pronounced modulations of the time-integrated FWM signal and of the unmodulated, broad temporal evolution of the time-resolved signal is only achieved when inhomogeneous broadening and many-body Coulomb effects [18] are taken into account. In particular, the screening of the Coulomb interaction has two important consequences for the FWM signals. First, it weakens the delayed interaction-induced part of the time-resolved signals and partly explains its featureless appearance. Secondly, the temporal position of the second maximum of the time-integrated signal, which is basically determined by the exciton binding energy, increases with increasing excitation density due to the decrease of the exciton binding energy.

This detailed analysis demonstrates that in general the real-time behaviour of the FWM signal is determined by a complicated interplay between dephasing, interaction-induced contributions, and inhomogeneous broadening. On the other hand, our analysis confirms that the temporal separation of the first and second maxima in the time-integrated curves as shown in, for example, figure 3 is indeed basically determined by the exciton binding energy. This implies in turn that transient FWM experiments on exciton wavepackets allow a determination of exciton binding energies, in particular in cases where this is difficult or impossible in linear spectroscopy, for example, because of considerable inhomogeneous broadening.

3.2. Exciton wavepackets in symmetrically strained (GaIn)As/Ga(PAs) multiple quantum wells: determination of exciton binding energies

Strained layer quantum wells and superlattice structures have become of considerable interest recently, because they provide an additional degree of freedom for tailoring the band structure properties; in particular the in-plane effective mass of the highest valence band states can be lowered by compressive strain [19]. Asymmetrically strained multi quantum wells and superlattices, like the (GaIn)As/GaAs structures discussed before, consist of alternating layers of strained (GaIn)As and unstrained GaAs material. In this case, the thickness of the individual layers as well as the total thickness of the multi quantum well or superlattice structure is limited to a critical

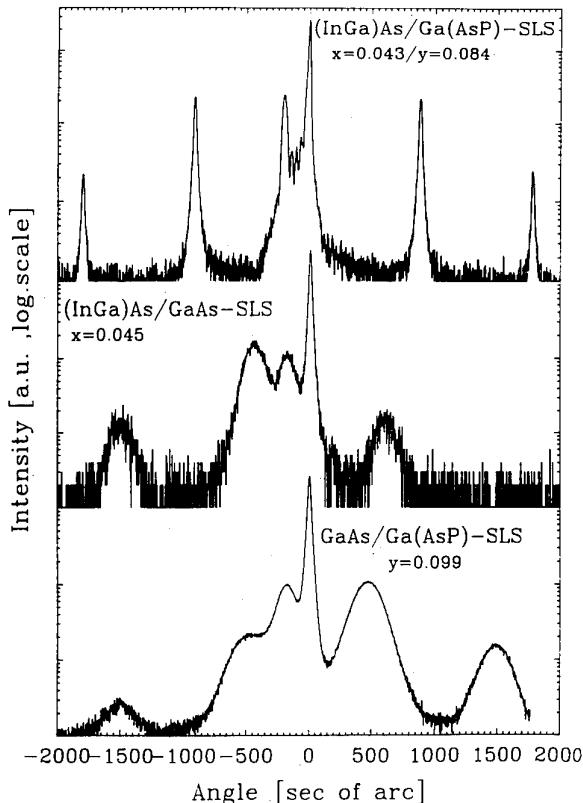


Figure 6. High-resolution x-ray diffraction pattern of a symmetrically strained $(\text{Ga}_{1-x}\text{In}_x)\text{As}/\text{Ga}(\text{P}_y\text{As}_{1-y})$ multi quantum well (upper curve) in comparison with asymmetrically strained structure (lower curves).

thickness where the pseudomorphic growth relaxes by forming dislocations. The limitation for the total thickness, however, can be overcome by growing symmetrically strained structures consisting of alternating layers with tensile and compressive strain. We have recently demonstrated that highly strained $(\text{Ga}_{1-x}\text{In}_x)\text{As}/\text{Ga}(\text{P}_y\text{As}_{1-y})$ multi quantum wells can be grown by MOVPE on GaAs with a total thickness of about 1 μm even for lattice mismatch ($\Delta a^\perp/a_0$) as large as 2.4% [20]. Nevertheless, these structures exhibit high-quality structural, electronic and optical properties comparable to unstrained material. The high structural quality is demonstrated in figure 6, where a high-resolution x-ray diffraction pattern ($\text{Cu K}_{\alpha 1}$ wavelength) around the (400) GaAs substrate reflection is shown for a symmetrically strained structure (upper curve) and compared with asymmetrically strained structures (lower curves). The upper curve exhibits narrow linewidth and pronounced satellite peaks with symmetric distribution around the central peak. The narrow peak at an angle of -200 arcsec corresponds to the $(\text{AlGa})\text{As}$ buffer layer grown on top of the GaAs substrate. The fact that only one further zeroth-order peak is observed implies that the zeroth-order peak of the multi quantum well sequence and the GaAs substrate lie on top of each other, demonstrating the perfect strain balance achieved. The linewidth of the individual peaks amounts to 20 arcsec which is very close to the theoretical limit of 18 arcsec, demonstrating that

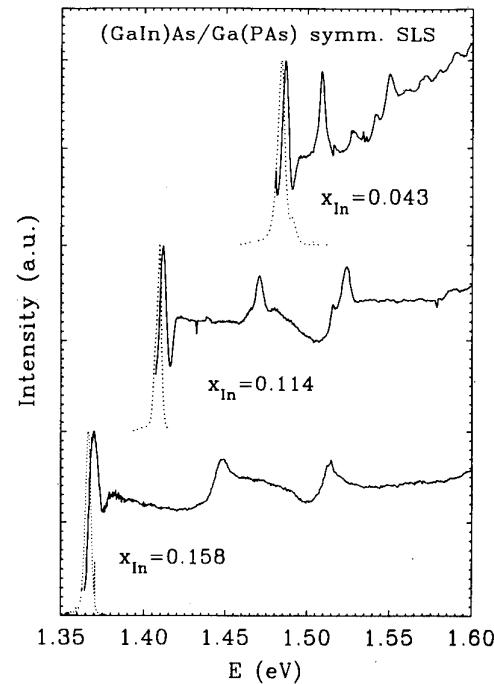


Figure 7. Photoluminescence excitation spectra (full curves) and photoluminescence spectra (dotted curves) at $T = 5$ K for symmetrically strained $(\text{GaIn})\text{As}/\text{Ga}(\text{PAs})$ multi quantum wells (for parameters see text) for three samples with different In content.

dislocation-free growth is achieved, independent of the In or P content within the range $x_{\text{In}} \leq 0.17$, $y_{\text{P}} \leq 0.35$. In contrast, the asymmetrically strained structures exhibit considerable broadening of the centre as well as the satellite peaks, indicating the poor structural quality.

The low-temperature PLE (full curves) and PL spectra (dotted curves) for symmetrically strained multi quantum-well structures consisting of 50 periods of 8.5 nm $(\text{GaIn})\text{As}/1.7$ nm GaAs/8.5 nm $\text{Ga}(\text{PAs})/1.7$ nm GaAs are shown in figure 7 for different In contents. With increasing In content a red-shift is observed for both the PL spectra and the PLE spectra. For the sample with $x_{\text{In}} = 0.171$ this red-shift amounts to 165 meV with respect to the GaAs bandgap. This red-shift reflects the formation of deeper quantum wells due to the decrease of the $(\text{GaIn})\text{As}$ bandgap with increasing In content. Accordingly, an increase of the exciton binding energy is expected in structures with high In content. However, in turn, the in-plane effective hole mass of the highest valence band subband is expected to decrease with In content due to the increase in built-in compressive strain [19], which would result in a decrease of the exciton binding energy. A precise determination of the exciton binding energy would thus provide a very powerful means of demonstrating that in fact the expected modification of the valence band structure is achieved. Unfortunately, a precise determination of the exciton binding energy from the PLE spectra is difficult for some samples, due to the inhomogeneous broadening which basically results from alloy disorder. We have therefore employed

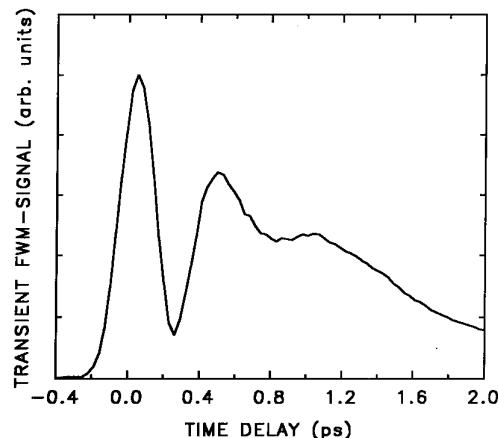


Figure 8. Time-integrated FWM signal at $T = 5$ K for a $(\text{Ga}_{0.906}\text{In}_{0.094})\text{As}/\text{Ga}(\text{P}_{0.22}\text{As}_{0.78})$ multi quantum well sample.

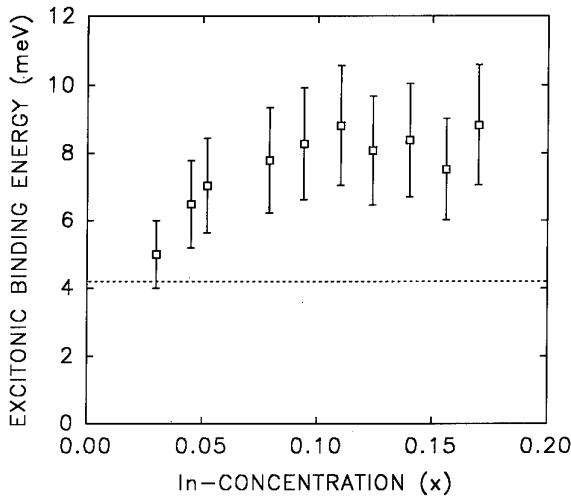


Figure 9. Exciton binding energies for symmetrically strained $(\text{GaIn})\text{As}/\text{Ga}(\text{PAs})$ multi quantum wells with different In concentration determined from the FWM data.

time-integrated FWM in order to study the dynamics of exciton wavepackets in these structures. As demonstrated above, this allows the determination of the exciton binding energy even in the presence of inhomogeneous broadening.

An experimental result of a transient FWM experiment with time-integrated detection of the nonlinear signal is shown in figure 8 for a sample with an In content of $x_{\text{In}} = 0.094$. The experimental results are very similar to those already discussed above for the $(\text{GaIn})\text{As}/\text{GaAs}$ structures; in particular the initial fast decay and subsequent recovery of the FWM signal reflecting the interference of the 1s and continuum exciton transitions are observed. The separation of the first and second maxima for this particular sample amounts to 500 fs corresponding to an exciton binding energy of 8.3 meV. For comparison, the exciton binding energy of bulk GaAs is 4.2 meV. The experimentally determined exciton binding energies for samples with different In content are plotted in figure 9. It is seen that only at low In

concentrations does the binding energy increase with increasing In concentration, which is attributed to the increasing electronic confinement. The saturation behaviour at higher In concentrations can be explained solely by the expected saturation of the confinement effect with increasing barrier height [21]. Hence, there is no indication of a pronounced decrease of the in-plane effective heavy-hole mass with increasing compressive strain in the $(\text{InGa})\text{As}$ layers.

4. Summary and conclusions

We have investigated the dynamics of excitonic wavepackets in $(\text{InGa})\text{As}/\text{GaAs}$ and symmetrically strained $(\text{InGa})\text{As}/\text{Ga}(\text{PAs})$ multi quantum well structures by transient FWM. We observe that the time-integrated FWM signal exhibits a pronounced modulation (beating) superimposed onto the overall decay of the nonlinear signal. This modulation reflects the coherent dynamics of exciton wavepackets. In particular, the pronounced initial minimum can be attributed to an interference between the 1s exciton transition and its continuum transitions. In contrast to the time-integrated FWM signals, no modulation is observed if the signals are time resolved. Instead, a broad echo-like response is found. A description of this behaviour is possible only by inclusion of inhomogeneous broadening and Coulomb interaction, meaning that the full semiconductor Bloch equations must be applied. Nevertheless, this analysis proves that the temporal separation of the first and second maxima in the time-integrated FWM traces basically corresponds to the inverse of the exciton binding energy. Thus, the investigation of the dynamics of exciton wavepackets can be applied for a determination of exciton binding energies. This is of particular importance for semiconductor nanostructures where optical transitions are often inhomogeneously broadened and a precise determination of the exciton binding energy is difficult.

We have applied this technique to determine the exciton binding energies in symmetrically strained pseudomorphic $(\text{InGa})\text{As}/\text{Ga}(\text{AsP})$ multi quantum wells. We find a saturation of the exciton binding energy with increasing In content. This effect can be attributed to the saturation of the confinement effect with increasing barrier height. There is no indication of a pronounced strain-induced decrease of the in-plane effective mass of the highest heavy-hole subband with increasing In concentration.

Acknowledgments

The expert technical assistance of M Preis and T Ochs is gratefully acknowledged. This work has been supported by the Deutsche Forschungsgemeinschaft through the Leibniz and Gerhard Hess Förderpreise.

References

[1] Leo K, Shah J, Göbel E O, Damen T C, Schmitt-Rink S, Schäfer W and Köhler K 1991 *Phys. Rev. Lett.* **66** 201

[2] Leo K, Göbel E O, Damen T C, Shah J, Schmitt-Rink S, Schäfer W, Müller J F, Köhler K and Ganser P 1991 *Phys. Rev. B* **44** 5726

[3] Leo K, Damen T C, Shah J, Göbel E O and Köhler K 1990 *Appl. Phys. Lett.* **57** 19

[4] Feuerbacher B F, Kuhl J, Eccleston R and Ploog K 1990 *Solid State Commun.* **74** 1279

[5] Feldmann J, Leo K, Shah J, Miller D A B, Cunningham J E, Meier T, von Plessen G, Schulze A, Thomas P and Schmitt-Rink S 1992 *Phys. Rev. B* **46** 7252

[6] Feldmann J 1992 *Festkörperprobleme/Advances in Solid State Physics* vol 32 ed U Rössler (Braunschweig: Vieweg) p 81

[7] Leo K, Bolivar P H, Brüggemann F, Schwedler R and Köhler K 1992 *Solid State Commun.* **84** 943

[8] Albrecht T F, Sandmann J H H, Feldmann J, Stolz W, Göbel E O, Nebel A, Fallnich C and Beigang R 1993 *Appl. Phys. Lett.* **63** 1945

[9] Bigot J Y, Mycek M A, Weiss S, Ulbrich R G and Chemla D S 1993 *Phys. Rev. Lett.* **70** 3307

[10] Lindberg M and Koch S W 1988 *Phys. Rev. B* **38** 3342

[11] Haug H and Koch S W 1993 *Quantum Theory of the Optical and Electronic Properties of Semiconductors* 2nd edn (Singapore: World Scientific); see also Schäfer W 1993 *Optics of Semiconductor Heterostructures* ed F Henneberger *et al* (Berlin: Akademie) p 21

[12] Leo K, Wegener M, Shah J, Chemla D S, Göbel E O, Damen T C, Schmitt-Rink S and Schäfer W 1990 *Phys. Rev. Lett.* **65** 1340

[13] Lindberg M, Binder R and Koch S W 1992 *Phys. Rev. A* **45** 1865

[14] Kim D S, Shah J, Damen T C, Schäfer W, Jahnke F, Schmitt-Rink S and Köhler K 1992 *Phys. Rev. Lett.* **69** 2725

[15] Weiss S, Mycek M A, Bigot J Y, Schmitt-Rink S and Chemla D 1992 *Phys. Rev. Lett.* **69** 2685

[16] Meier T 1994 *PhD Thesis* Marburg

[17] Feldmann J, Meier T, von Plessen G, Koch M, Göbel E O, Thomas P, Bacher G, Hartmann C, Schweizer H, Schäfer W and Nickel H 1993 *Phys. Rev. Lett.* **70** 3027

[18] Wang H, Ferrio K, Steel D G, Hu Y Z, Binder R and Koch S W 1993 *Phys. Rev. Lett.* **71** 1261

[19] Houg S, Jaffe M and Singh J 1987 *IEEE J. Quantum Electron.* **23** 2181

[20] Lutgen S, Marschner T, Albrecht T F, Stolz W and Göbel E O 1993 *Mater. Sci. Technol.* **B 21** 249

O'Reilly E P 1989 *Semicond. Sci. Technol.* **4** 121

[21] Koch M, Volk M, Meier T, Feldmann J, Stolz W, Thomas P and Göbel E O 1994 *Superlatt. Microstruct.* at press

Picosecond and femtosecond spectroscopy of highly excited charge carriers in semiconductors: theory

P Kocevar

Karl-Franzens-Universität Graz, Institut für Theoretische Physik,
Universitätsplatz 5, A-8010 Graz, Austria

Abstract. Time-resolved luminescence, absorption and four-wave-mixing spectroscopy in the picosecond and femtosecond regimes are currently the most efficient experimental methods for studying energy and phase relaxation rates of highly excited charge carriers in solids. Here most of the effort concerns semiconductors, as these relaxation rates will determine and limit the efficiency of future generations of ultrafast electronic and optoelectronic devices. The following survey describes some attempts at a detailed analysis of recent experimental luminescence and absorption data covering a representative range of materials, spectral regimes and excitation densities. In spite of the fact that nanostructures are at the focus of present-day technological developments our discussion is limited to experiments on unstructured semiconductors, because the much simpler electronic structure of bulk materials will make it easier to realize the essentials of an extremely transient generation and relaxation dynamics of charge carriers and its description by appropriate Monte Carlo or quantum transport techniques.

1. Introduction

The last decade has seen the rapid development of pulsed-laser technologies towards higher intensities and shorter pulse durations with spectroscopic time resolutions finally approaching the femtosecond regime. These developments have brought a merging of two originally well-separated fields of semiconductor research, both concerned with the nonlinear electronic response to high photoexcitation.

Since the early 1970s the hot-electron community has been investigating the nonlinear electrical (i.e. non-ohmic) response of free charge carriers to strong DC electric fields. The prospect of ultrafast semiconductor devices, including optoelectronic switches and sensors, and the above-mentioned advent of fast laser spectroscopy then prompted the widespread use of time-resolved optical techniques to study the energy relaxation of ‘hot’ carriers in the picosecond and subpicosecond time regimes. The reason for this change from transport to optics was obvious. A nonlinear DC transport coefficient unavoidably results from a very involved interplay of many relevant dissipation channels, very often summarized in oversimplifying global relaxation rates for momentum, energy and total particle number. An appropriately designed energy-loss spectroscopy of photoexcited carriers could provide much complementary information on individual dissipation channels and scattering mechanisms. Here the most direct insights have been obtained by time-resolved luminescence and

absorption (i.e. transmission/reflection) measurements [1–6]. Moreover, similar time-resolved spectroscopies have permitted study of additional and more subtle relaxation phenomena such as orientational relaxation [7, 8] and spectral hole burning [1, 9].

Besides the mainly device-oriented hot-carrier community a different group of researchers has for many years been very actively investigating the coherent interaction of electrons and a laser field by studying the nonlinear optical response of atoms or molecules and of excitons in solids, including bulk and microstructured semiconductors [10–15].

Naturally there was a lot of basic scientific interest in detecting and theoretically describing coherence phenomena of *localized* carriers such as the AC Stark effect, Rabi oscillations of photoexcited state populations, quantum beats between optically prepared coherent states, photon echoes or self-induced transparency. However, a further stimulus was the technological prospect of practical applications in future generations of ultrafast electro-optic or all-optic devices by utilizing analogous effects in the optical response of *free* carriers in some of the standard materials of present-day semiconductor technology. And such effects were indeed observed as early as 1988 in a pioneering experiment on coherent transients of free carriers and their dephasing rates in the few-femtoseconds regime [16].

In semiconductors (or semiconductor nanostructures) even the shortest laser pulses can create up to several 10^{18} per cm^3 (or 10^{11} per cm^2) photoexcited

carriers [16–18]. Depending on the intensity, spectral distribution and temporal shape of the excitation pulse and also on the time resolution in the detection of the optical response, the underlying physics shows or promises a wide range of interesting phenomena.

All excitation scenarios of our following examples belong to the non-excitonic regime of moderate or high excitation energies and densities for which the electronic response is completely governed by the dynamics of free carriers. Here a very decisive role is played by the leading-order incoherent processes, i.e. by the carrier–carrier (cc) and carrier–phonon (cph) scatterings, the latter mainly involving near-zone-centre longitudinal (LO) or transverse (TO) optical phonons and, for electronic transfers between conduction-band valleys, phonons from near-zone-boundary modes.

2. Incoherent nonlinear optical response of free carriers

2.1. Semiclassical kinetic transport description: ensemble Monte Carlo simulations

The details of the scattering dynamics are decisive for the understanding of ‘hot’ luminescence and of state-filling effects in transient absorption spectra, because the respective nonlinearities are, apart from density-of-states factors and suitably averaged optical transition amplitudes, proportional to the product or the sum of the carrier distribution functions for the optically coupled states in each contributing detection channel.

The best way to describe the scattering dynamics within a non-equilibrium carrier system in terms of single-particle phase-space distribution functions is the time-dependent Boltzmann equation (TBE) with its adaptions to solid state applications through use of quantum-mechanical scattering cross sections and rates for generation, trapping and recombination of carriers, and with occasional completion by analogous kinetic equations for non-equilibrium phonon distributions [19]. One should nevertheless be aware that the Boltzmann picture of ballistic particles on free trajectories, interrupted by instantaneous scattering events with uncorrelated initial states, is pushed towards its limits when used to describe dense carrier systems with strong Coulombic cc and cph interactions.

In our following examples the TBE was solved by ensemble Monte Carlo (EMC) simulations of the coupled carrier–phonon system in k space. A typical EMC simulation follows the time evolution of about 10 000 particles as representatives of the actual carrier system by using random numbers in combination with given probability distributions for photoexcitation rates and total and differential scattering cross sections [20]. While including all details of the band structure in the photogeneration process, all the (secondary) effects of band non-parabolicities and anisotropies on the scattering dynamics were neglected. The spectral and temporal profile of excitation, probe and luminescence up-conver-

sion pulses were accounted for by correspondingly structured eh (electron–hole) pair generation rates and by appropriate frequency and time convolutions of the calculated optical response. All relevant cc and cph scatterings were taken into account. Carrier degeneracy and, if necessary, non-equilibrium phonons were included by use of suitable rejection techniques [3]. Recombination processes and acoustic-phonon scatterings were neglected as they would not contribute at the short time-scales of our present interest.

For the cc interaction and the long-range polar optic coupling of electrons (e) and holes (h) to LO phonons (LOP) in polar materials a time-dependent, self-consistent screening model was used, in which the EMC simulation easily allows the evaluation of the long-wavelength static limit of the RPA dielectric function for arbitrary non-equilibrium carrier distribution functions at each time instant. To provide for the strong inelasticity of the cc and c–LOP scatterings a further dynamical ingredient was included by mass selectivity, allowing only sufficiently light particles to screen a given scattering event [3].

2.2. Applications

As our first example we present the preliminary results of an EMC analysis [21] of the recent 2 ps pump–2 ps probe infrared (IR) absorption spectroscopy of heavy (hh) and light holes (lh) in p-doped Ge [5]. Such studies have some relevance in connection with the finding of various external-field-induced stimulated IR emissions in this material. Here a long-standing discussion exists about whether IR excitation with moderate or high excitation densities could lead to pronounced state-filling effects in the lh band or just to a very fast internal thermalization of light and heavy holes, followed by ‘conventional’ cooling dynamics of the (hh + lh) carrier plasma [5]. Our results show that both phenomena indeed contribute to the picosecond time evolution of the spectra. Figure 1 shows the experimental and theoretical results for the pump-pulse-induced relative absorption change at the excitation frequency together with the simulated hh and lh distribution functions as functions of time. As in all the following examples the most important experimental and material parameters (in standard notation) are listed in the figures, the zero of time taken at the centre of the excitation pulse. For both the acceptor–lh and the hh–lh absorption channels (indicated by the vertical lines in the lower figure) the initial bleaching is seen to be caused by a strongly athermal enhancement of the final-state occupation in the lh band and a corresponding dip in the initial states of the hh–lh absorption channel, in accordance with the non-thermal-distribution argument. However, the ensuing absorption gain after the end of the excitation pulse is then indeed in agreement with the standard hot-carrier interpretation, as it is caused by the replenishment and thermalization of the hh band population, resulting in an increase of the initial-state hh distribution function, with negligible lh occupations due to the very rapid lh → hh transfers via cph and cc scatterings.

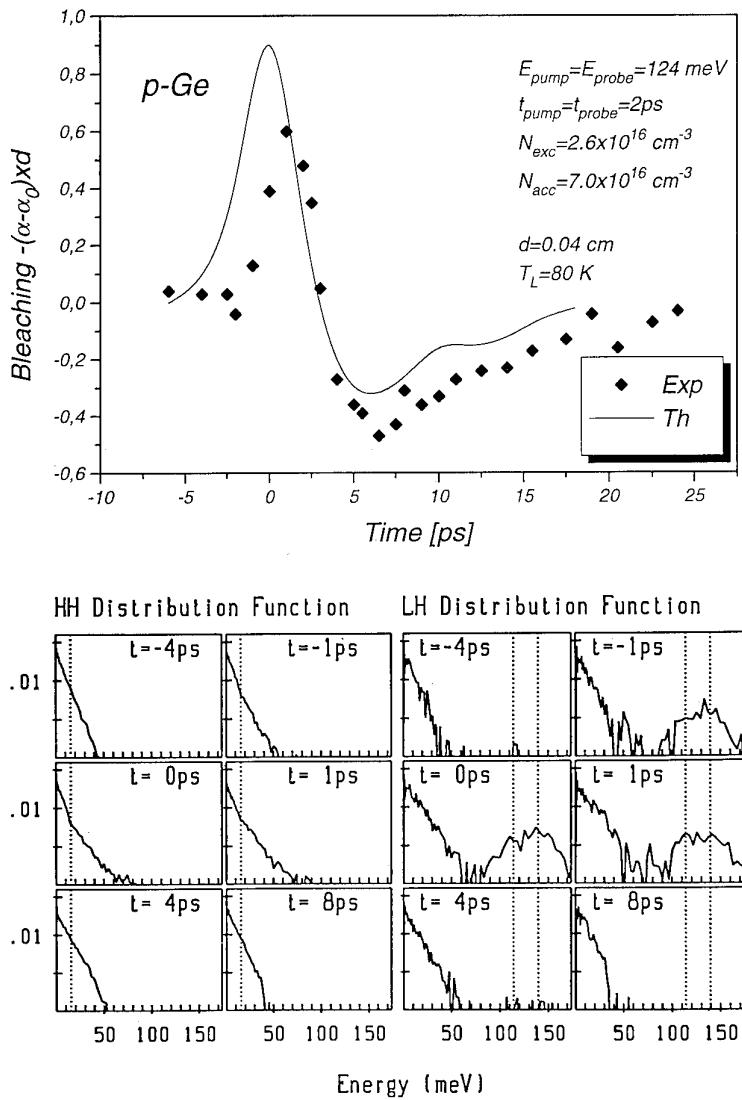


Figure 1. Upper part: change of IR absorption versus time; lower part: distribution functions of holes at different times.

Our second example concerns the first direct experimental demonstration [22] of the so-called 'hot-phonon' effect. This picosecond phenomenon has for many years been theoretically predicted and is caused by an amplification of strongly coupling optical-phonon modes through the pronounced phonon emission by highly photoexcited carriers in the early relaxation stage during and after a laser pulse excitation. This leads to strongly increased rates for phonon reabsorption by low-energy carriers and to a resulting reduction of the energy relaxation of the total carrier plasma.

Figure 2 shows the so-called cooling curves obtained through an effective temperature parameter from the near-exponential frequency dependence of the luminescence intensities in a two-pulse experiment using 80 fs laser pulses with photon energies of 1.72 eV, exciting e-h pairs across the bandgap of 'low-temperature' GaAs. This material is especially prepared to contain very effective free-carrier traps, resulting in carrier lifetimes of about 1 ps, several times smaller than the LOP lifetime.

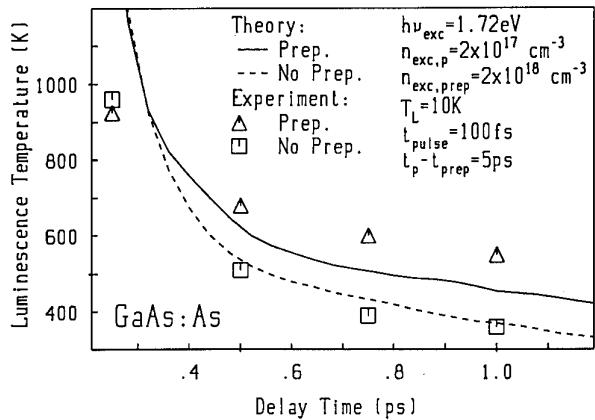


Figure 2. Plasma cooling with and without prepump.

Consequently the carriers excited by a strong prepump pulse will disappear after a few picoseconds, leaving a still pronounced non-equilibrium LOP population. This

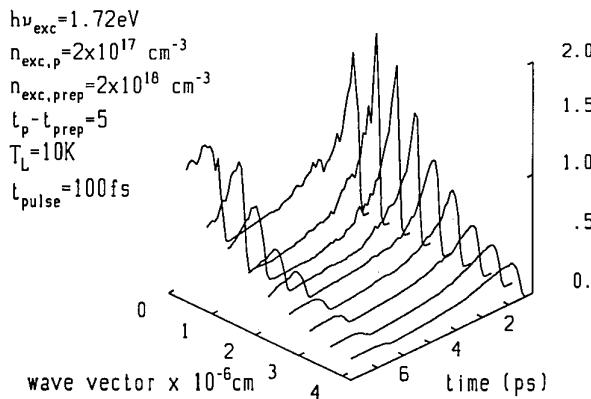


Figure 3. LO phonon amplification with prepump.

situation is then probed by a second, much weaker pump pulse, generating an e-h density sufficiently small to preclude a noticeable free-carrier screening of the polar-optic c-LOP interaction. This then allows a direct study of the effect of LOP disturbances on the cooling of the second carrier generation by varying the time delay between the two pulses and the intensity of the prepump pulse. By use of the experimentally known LOP lifetime for decays into pairs of phonons belonging to (electronically inactive) near-zone-boundary modes the simulations at least semiquantitatively reproduce the experimental data, as shown in figure 2. The corresponding evolution of the LOP distribution function is seen in figure 3, demonstrating the pronounced optical-phonon amplification in this type of laser-pulse excitation as exemplified for many years in the time-resolved Raman spectroscopy of optical phonons.

3. Coherent nonlinear optical response of free carriers

3.1. Quantum density matrix description

The dominance of the phase-breaking scattering processes under moderate and high excitation conditions raises the question whether coherent transients, and in particular whether Rabi-type oscillations in the time-resolved absorption, might be directly detectable in the nonlinear spectroscopy of free carriers in semiconductors.

The conventional theoretical framework for describing the nonlinear coherent response to laser excitation has been the semiconductor Bloch equations (SBE). These are the generalization of the well-known Bloch equations for a single optically driven two-level oscillator to an ensemble of oscillators, each oscillator consisting of an optically coupled pair of electron and hole states at a given wavevector k . The system response is then characterized by the respective occupation numbers f_k^e and f_k^h and by the e-h pair amplitudes p_k , where the p_k are directly related to the induced polarization and thereby to the optical susceptibility.

In addition to the laser-field-induced renormalizations of and couplings between these oscillators there are additional contributions from the cc and cph interactions.

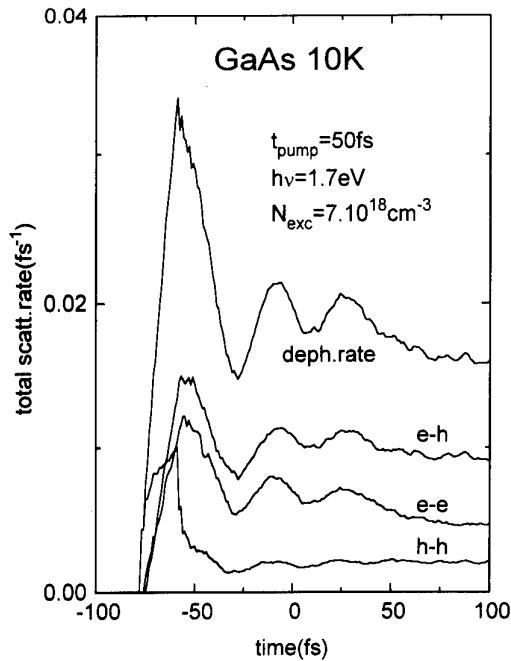


Figure 4. Dephasing and cc scattering rates versus time.

They have in most applications been treated through phenomenological damping rates in terms of a constant phase relaxation time T_2 for the e-h amplitude and a relaxation time T_1 describing the changes of the individual state occupations due to scatterings. To be sure, such a 'T'ology is handy for a qualitative interpretation of experimental data, but will never provide a quantitative description.

As illustration of the need to go beyond the use of constant relaxation times figure 4 shows the simulated ee, eh, hh and total cc scattering rates for a 1.72 eV excitation scenario in GaAs, which will serve as the model case in our following analysis. It is seen that the total cc scattering rate γ_{cc} , which practically equals the dephasing rate, sharply increases during the build-up of the eh density and thereafter strongly decreases due to the cooling-induced increase of the free-carrier screening of the cc interaction. The replacement of γ_{cc} by a constant $1/T_2$, which is usually adjusted to some average cc rate, will strongly underestimate the dephasing during and immediately after the pump pulse. A similar oversimplification would result during periods of rapidly changing contributions of other non-negligible dissipation channels, as, for example, during the onset of intervalley or interband repopulations or of optical-phonon amplification.

So we finally turn to some recent attempts at a consistent treatment of the scatterings within a first-principles quantum formulation of the coherent carrier dynamics [23, 24]. Following earlier such formulations [25–27], we have, in collaboration with T Kuhn and F Rossi, generalized the first successful implementation of EMC simulations into first-principles formulations [27] to the cases of arbitrary complex excitation and scattering

scenarios typical for a high-excitation femtosecond spectroscopy of semiconductors.

For that purpose the exact equations of motion (EOM) for the two- and four-point and for the phonon-assisted three-point density matrices in k -space (taken as diagonal, i.e. assuming spatial homogeneity) are derived within the Heisenberg picture up to second order in the cc and cph couplings, and the resulting hierarchy is truncated by suitable factorizations of the higher-order correlation functions. The usual adiabatic and Markov approximations for the three- and four-point functions then lead to a closed set of EOM for the e-e, h-h and e-h density matrices, that is for f_k^e , f_k^h and p_k .

Let us, for simplicity, restrict our discussion to the cc interaction. The contributions of first order in the (consistently screened) cc potential V_q are still fully coherent, leading to a mean-field (i.e. Hartree-Fock) renormalization of the single-particle energies and of the electric field. They thereby modify the Rabi frequency of the initially uncoupled Bloch oscillators and, more importantly, add new coherence-driven e-h generation and recombination processes. The leading incoherent (i.e. scattering) contributions appear in second order. In the EOM for f_k^e and f_k^h they (naturally) are the familiar Boltzmann collision integrals, with very similar expressions for the scattering contributions to p_k .

The great practical advantage of this approach is the fact that, after implementing the various renormalizations and coherence-induced modifications of the photo-generation into the EMC algorithm, the EOM for p_k can be integrated in parallel with the generalized EMC simulation of the e-h system. In this way one solves the complete set of EOM consistently up to second order in V_q and, as shown earlier [27], also in the cph couplings. As p_k is directly related to the internal polarization, one thereby obtains the optical susceptibility and the differential absorption constant as functions of time. At last this ‘bare’ absorption spectrum has to be folded with the spectral and temporal shape of the probe pulse to conform with the experimental detection procedure.

3.2. Applications

For simplicity the following results were obtained for a two-band model of GaAs; in this way the initial conduction-band occupations are concentrated into one 1.7 eV excitation and detection channel (hh-e) instead of two (hh-e, lh-e) channels, resulting in a more pronounced initial state filling and a corresponding overestimate of Rabi oscillation amplitudes.

For a 50 fs pump pulse its intensity I_{pump} was chosen to correspond to an unrenormalized Rabi period $t_R^{(0)}$ ($\propto I_{\text{pump}}^{-1/2}$) of 33 fs. Figure 5 shows the net generation and recombination rates. It is seen that at three time intervals recombination outweighs generation, leading to a non-monotonic behaviour of the e-h density and to corresponding Rabi-type ‘wiggles’. Even after the broadening through the time and frequency foldings with respect to a 15 fs probe pulse these oscillatory features are still visible in the relative absorption changes depicted in

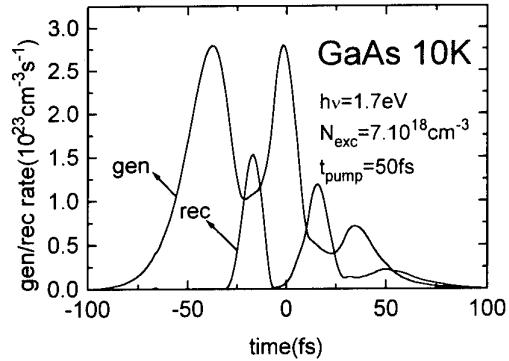


Figure 5. Generation/recombination rates versus time.

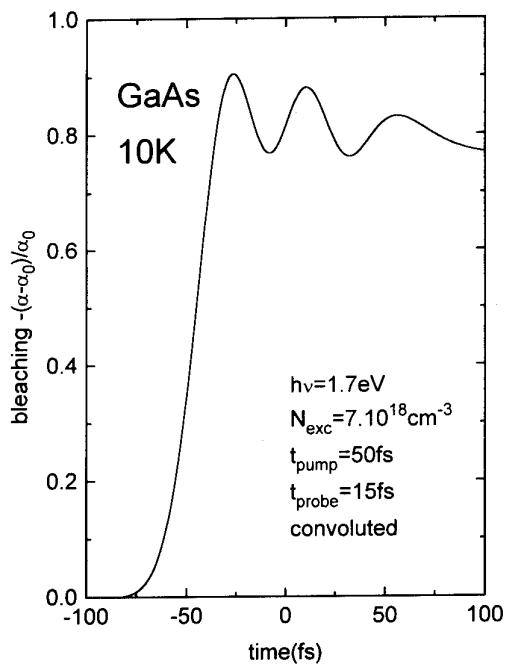


Figure 6. Convolved bleaching versus time.

figure 6; they develop at least three Rabi oscillation periods before being damped out by the dephasing action of the cc (and to a minor extent of the cph) scatterings.

This prediction of a favourable scenario for the detection of transient coherence effects in highly excited semiconductors of course mainly depends on the details of the theoretical description, and here in particular on the quality of our screening model for the phase-breaking cc scatterings. Various research groups are currently working on an improved non-equilibrium screening theory, in view of the unsatisfactory *ad hoc* use of a screened cc potential in the Hamiltonians underlying our present and various related many-body calculations (which ansatz more or less corresponds to the ‘screened exchange’ approaches for equilibrium systems). In any case a numerically tractable improvement will have to be found in view of the complicated microscopic scenarios in this type of nonlinear physics.

References

[1] Lin W Z, Schoenlein R W, Fujimoto J G and Ippen E P 1988 *IEEE J. Quantum Electron.* **24** 267

[2] Rota L, Lugli P, Elsaesser T and Shah J 1993 *Phys. Rev. B* **47** 4226

[3] Hohenester U, Supancic P, Kocevar P, Zhou X Q, Kütt W and Kurz H 1993 *Phys. Rev. B* **47** 13233

[4] Snoke D W, Röhle W W, Lu Y-C and Bauser E 1992 *Phys. Rev. B* **45** 10979

[5] Woerner M, Elsaesser T and Kaiser W 1992 *Phys. Rev. B* **45** 8378

[6] Elsaesser T, Woerner M, Portella M T, Frey W, Ludwig C and Lohner A 1994 *Semicond. Sci. Technol.* **9** 689

[7] Smirl A L, Boggess T F, Wherrett B S, Perryman G P and Miller A 1983 *IEEE J. Quantum Electron.* **19** 690

[8] Oudar J L, Migus A, Hulin D, Grillon G, Etchpare J and Antonetti A 1984 *Phys. Rev. Lett.* **53** 384

[9] Oudar J L, Hulin D, Migus A, Antonetti A and Alexandre F 1985 *Phys. Rev. Lett.* **55** 2074

[10] Mysorowicz A, Hulin D, Antonetti A, Migus A, Masselink W T and Morkoc H 1986 *Phys. Rev. Lett.* **56** 2748

[11] Von Lehmen A, Chemla D S, Zinker G E and Heritage G P 1986 *Opt. Lett.* **11** 609

[12] Fluegel B, Peyghambarian N, Olbright G, Lindberg M, Koch S, Joffre M, Hulin D, Migus A and Antonetti A 1987 *Phys. Rev. Lett.* **59** 2588

[13] Honold A, Schultheis L, Kuhl J and Tu C W 1989 *Phys. Rev. B* **40** 6442

[14] Langer V, Stoltz H and von der Osten W 1990 *Phys. Rev. Lett.* **64** 854

[15] Leo K, Wegener M, Shah J, Chemla D S, Goebel E O, Damen T C, Schmitt-Rink S and Schäfer W 1990 *Phys. Rev. Lett.* **65** 1340

[16] Becker P C, Fragnito H L, Brito Cruz C H, Fork R L, Cunningham J E, Henry J E and Shank C V 1988 *Phys. Rev. Lett.* **61** 1647

[17] Knox W H, Chemla D S, Livescu G, Cunningham J E and Henry J E 1988 *Phys. Rev. Lett.* **61** 1290

[18] Goodnick S M, Lugli P, Knox W H and Chemla D S 1989 *Solid-State Electron.* **32** 1737

[19] Peierls R 1955 *Quantum Theory of Solids* (London: Oxford University Press)

[20] Jacoboni C and Lugli P 1989 *The Monte Carlo Method for Semiconductor Device Simulation* (Berlin: Springer)

[21] Schullatz M 1993 *Diploma Thesis* University of Graz

[22] Zhou X Q, Röhle W W, Schullatz P, Kocevar P, van Driel H M and Ploog K 1994 *Semicond. Sci. Technol.* **9** 704

[23] Adler F J, Kocevar P, Schilp J, Kuhn T and Rossi F 1994 *Semicond. Sci. Technol.* **9** 446

[24] Adler F J, Kuras G F and Kocevar P 1994 *SPIE Proc.* **2142** at press

[25] Lindberg M and Koch S W 1988 *Phys. Rev. B* **38** 3342

[26] Kuznetsov A V 1991 *Phys. Rev. B* **44** 8721

[27] Kuhn T and Rossi F 1992 *Phys. Rev. B* **46** 7496

Theory of ultrafast phenomena in laser-excited semiconductors

D K Ferry

Arizona State University, Tempe, AZ 85287-5706, USA

Abstract. The study of semiconductor carriers excited by femtosecond laser excitation has proved to be one of the most effective methods of characterization of semiconductors, as well as providing a probe for new physical effects. On the femtosecond time-scale, however, it is not expected that the semiclassical Boltzmann equation will be correct, and a more basic quantum mechanical formulation is required. The formulation of equivalent transport equations from the Schrödinger equation and its variants will be briefly reviewed. Results, using an ensemble Monte Carlo method, for electron-electron and electron-phonon interactions will be presented to show how these studies can be used to investigate the strength of the electron-phonon interaction and how the finite collision duration affects such studies.

1. Introduction

With the push for smaller device dimensions and higher-frequency operation, hot-electron phenomena, particularly ultrafast scattering processes on the femtosecond time-scale, have become an extensive area of research. A number of groups have been active in studying the intervalley scattering among the Γ , L and X valleys in GaAs [1–6], which is one of the dominant scattering mechanisms for hot electrons with energies above the intervalley scattering threshold. Experimental techniques based on femtosecond optical excitation and measurements and transport are the basis for various determinations of the intervalley scattering parameters through numerical fit to measurements of population evolutions monitored by pulsed photoexcitation pump-probe techniques [1, 5], linewidth broadening of the photoluminescence spectra in continuous-wave optical excitation [2, 3] and mobility measurements [7, 8]. Numerical simulations using ensemble Monte Carlo (EMC) techniques are generally employed for parameter fitting to such data.

On the femtosecond time-scale, however, it is not expected that the semiclassical Boltzmann equation will be correct [9], and a more basic quantum mechanical formulation is required. In this paper, the formulation of equivalent transport equations from the Schrödinger equation and its variants will be briefly reviewed, to show how a Boltzmann limit may be reached. Equations on the femtosecond time-scale involve the presence of the transient polarization as well, and the higher-order interactions of this process are connected with equivalent many-body effects in nanostructures and mesoscopic systems. Results, using an ensemble Monte Carlo method, for electron-electron and electron-phonon interactions will be presented to show how these studies can be used to investigate the strength of the electron-

phonon interaction and how the finite collision duration affects such studies.

2. From Schrödinger to Boltzmann

Most approaches to quantum transport begin with the Schrödinger equation. In general one can solve the Schrödinger equation by assuming an expansion of the wavefunction, in some useful basis, and then developing an integral solution, in terms of an initial condition, as

$$\Psi(\mathbf{r}, t) = \int d\mathbf{r}' K(\mathbf{r}, t; \mathbf{r}', t_0) \Psi(\mathbf{r}', t_0) \quad (1)$$

where K is the kernel, or propagator. In general, this can be ensemble averaged over some distribution, or in the present case, over the basis states selected. The generalized propagator can then be defined as the Green function [10]. For example, the retarded Green function is given by

$$G_r(\mathbf{r}, \mathbf{r}'; t, t') = -i\theta(t - t') \langle \{\Psi(\mathbf{r}, t), \Psi^+(\mathbf{r}', t')\} \rangle. \quad (2)$$

In general, we must deal with a family of Green functions. Usually one only has to worry about the retarded and advanced functions, but here the distribution function is no longer the equilibrium Fermi-Dirac distribution. Consequently, we must also determine an appropriate distribution function. The retarded and advanced functions define the spectral density function, which relates the momentum to the energy. In semiclassical transport this is just a delta function, but in quantum transport this is no longer the case [10]. The distribution function is related to the so-called ‘less than’ Green function. Moreover, in the laser excitation area we must also worry about both electrons and holes, so that one wavefunction in (2) may represent a hole wavefunction and one an electron wavefunction. This complicates the problem by

now giving us 16 possible Green functions; fortunately they are not all independent of one another [11].

The ‘less than’ Green function is a two-time correlation function describing both position and temporal correlations. If we make the two time arguments equal, the result is simply the density matrix [12]. If we further make the two space arguments equal, then we just get the probability density which, within limits and constants, is related to the spatial charge density [13]. Thus, one can achieve a limiting transition to a function which can be related to the classical distribution function. This connection has allowed much work in the area of quantum transport for device modelling [13].

If the system Hamiltonian is composed of terms for the kinetic energy of electrons and holes, phonon terms, and the laser interaction term (in terms of the *field* of the laser), then the quantum equations of motion for the Green function, in the limit of equal times and positions and no scattering, lead to [14]

$$\begin{aligned} \frac{df_k^e}{dt} &= \frac{df_{-k}^h}{dt} = g_k^0(t) \\ &= \frac{2\pi}{ih} [\mathbf{M}_k \cdot \mathbf{F}_0(t) \exp(-i\omega_L t) p_k^* \\ &\quad - \mathbf{M}_{-k}^* \cdot \mathbf{F}_0^*(t) \exp(i\omega_L t) p_k] \\ \frac{dp_k}{dt} &= \frac{2\pi}{ih} \left[(E_k^e + E_{-k}^h) p_k + \mathbf{M}_k \cdot \mathbf{F}_0(t) \right. \\ &\quad \times \left. \exp(-i\omega_L t) (1 - f_k^e - f_{-k}^h) \right]. \end{aligned} \quad (3)$$

Here, E_k is the kinetic energy (for the electron or hole), f_k is the distribution function, p_k is the polarization in the electron–hole gas, \mathbf{F}_0 is the laser electric field and \mathbf{M}_k is the dipole matrix element connecting the hole state with the electron state. The distribution functions and polarization are the direct limits of the appropriate Green functions. If the polarization rapidly follows the excitation (ignoring the left-hand side of the last equation), then a Boltzmann equation results for the excitation and streaming terms. The equations (3) are often termed the semiconductor Bloch equations [14–16].

Dissipative processes can be added to (3) quite easily, but must be done with care. The interaction with the phonons is directly added to the Hamiltonian, and therefore to the equations for the electron and hole distribution functions through self-energy terms, for which the weak scattering process produces the equivalent semiclassical results that have been used for some time. The difference here is that the distribution functions in (3) are the limits of properly renormalized Green functions, so that care must be taken in the dissipative process to ensure that the renormalized distributions are used accordingly. In general, the limiting processes on both time and position must be done after formulation of the equations of motion and do not directly lead to the Boltzmann equation [17]. So, in this sense, while the formulation is the same as the Fermi golden rule used

semiclassically, the basis is much deeper. However, this means that semiclassical techniques can be suitably modified to probe into the corrections that quantum effects will play in the transport.

The other major corrections are the appearance of the polarization of the electron–hole gas and the possibility that the collisions are sufficiently strong that the collisional broadening will affect the local density of states. The former is most important for the very short pulse case, while the latter is not particularly important in high-quality semiconductor material, except for the role of carrier–carrier scattering in high excitation densities. We address this below.

3. The electron–electron interaction with exchange

Treatment of electron–electron interaction in solids analytically has always involved the need to approximate the full frequency- and wavevector-dependent dielectric function. This is difficult in femtosecond laser excitation due to the high temporal frequencies involved. When the carrier–carrier interaction is added to the Hamiltonian (in a momentum-state basis), the equations of motion lead to three-, four- and six-point Green functions in the equations for the two-point Green functions that give rise to (3). One often makes an unwarranted assumption that these can be factored into products of two-point functions [12]. However, this ignores correlations that still remain, and which are still multipoint functions [18–20]. These can give rise to long-lived correlations that are solved from, for example, the Bethe–Salpeter equation, and are known to contribute to DC quantities such as universal conductance fluctuations and weak localization. These terms have not been investigated in the laser case, and ignoring them is unwarranted at this time. Even when these correlations are ignored, such as when an assumption of adiabaticity and Markovian behaviour are made, the use of simply a frequency- and momentum-dependent RPA screening term neglects plasmon interactions, which are known to be important on the short-time basis [21]. One way to avoid these approximations is to treat a numerical simulation of the carriers and to incorporate the full carrier–carrier interaction, without approximations, in real space. Here the forces between the individual particles are calculated from direct Coulomb forces between each pair of carriers.

Consider an ensemble of electrons in which normal scattering and transport are treated by an ensemble Monte Carlo calculation. Intercarrier Coulomb interaction is retained as a real-space potential, and the effect of the interparticle potential on the motion of the electrons is computed through a molecular dynamics procedure in which the local force on each electron due to the electric field and the repulsion of all other electrons is calculated for each time step of the Monte Carlo process [22, 23]. Only a finite number of particles can be treated, so that a small cell of real space is considered and assumed to be replicated throughout the entire crystal [24].

At densities near 10^{18} cm^{-3} , which can now be easily obtained, one must begin to incorporate exchange contributions to the Coulomb energy. These contributions are directly a quantum mechanical modification of the semiclassical Boltzmann equation and the classical interaction treated by Coulomb forces [25]. The exchange interaction modifies the Coulomb interaction between the carriers by changing both the force itself and by an effective mass change. These changes result in the Hamiltonian equations

$$\begin{aligned}\frac{dp_i}{dt} &= -\frac{\partial H}{\partial x_i} = \sum_{j \neq i} F_{ij}^D + \sum_{j \neq i} \delta_{\sigma,\sigma_j} F_{ij}^{XC} \\ \frac{dx_i}{dt} &= \frac{\partial H}{\partial p_i} = \frac{p_i}{m} - \sum_{j \neq i} \delta_{\sigma,\sigma_j} G_{ij}\end{aligned}\quad (4)$$

where

$$\begin{aligned}F_{ij}^{XC} &= V_\sigma(\mathbf{k}_j - \mathbf{k}_i) \frac{\partial \Delta_\sigma(\mathbf{x}_j - \mathbf{x}_i)}{\partial x_i} \\ G_{ij} &= -G_{ji} = \frac{2\pi}{h} \frac{\partial V_\sigma(\mathbf{k}_j - \mathbf{k}_i)}{\partial \mathbf{k}_i} \Delta_\sigma(\mathbf{x}_j - \mathbf{x}_i)\end{aligned}\quad (5)$$

and

$$\begin{aligned}\Delta_\sigma(\mathbf{x}) &= \frac{1}{(4\pi\sigma^2)^{3/2}} \exp\left(-\frac{x^2}{4\sigma^2}\right) \\ V_\sigma(\mathbf{k}) &= \int d^3x V(x) \exp\left(i\mathbf{k} \cdot \mathbf{x} - \frac{x^2}{4\sigma^2}\right).\end{aligned}\quad (6)$$

The latter two functions are the transition from exchange to direct weighting function, and the generalized Fourier transform of the direct Coulomb force respectively. Now, however, it is necessary to compute a value for the quantity σ . For this, we evaluate the total exchange energy and use a value of s that makes this latter quantity equal to the homogeneous value in the momentum representation [25].

In figure 1, the relaxation time constant for an electron-hole plasma produced by a femtosecond laser excitation is shown. The data are compared with lifetimes

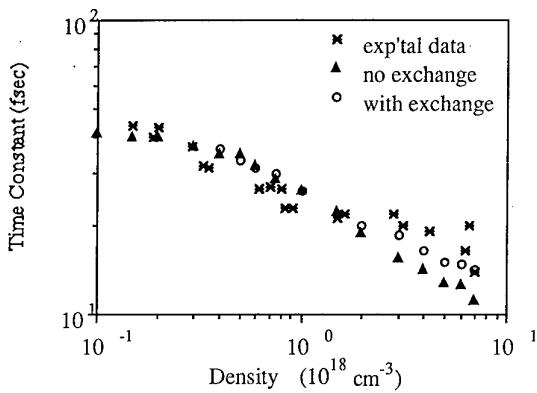


Figure 1. The relaxation time for a femtosecond laser excited plasma as calculated with a coupled ensemble Monte Carlo and molecular dynamics approach. The experimental data are taken from [26].

inferred from photon-echo experiments [26]. The time constants obtained from the experiment agree well with the calculations done through the coupled EMC/MD process, but there is still some disagreement at high density, which may be due to the neglect of the variations in the response caused by the electron-hole polarization.

4. Finite collision duration

As we discussed above, the use of the Fermi golden rule for electron-phonon scattering is an approximation only valid in the strict semiclassical limit. On the femtosecond time-scale, it is unlikely that a 30 meV phonon, which corresponds to a time period of (for the Fourier frequency of the phonon) 140 fs, can be fully emitted in only 10–20 fs. When the electron first begins to interact with the lattice, all phonon modes and frequencies are involved. Only in the long-time limit of the Fermi golden rule is a single frequency and mode finally involved. Thus, we have to pursue a more rigorous study of the electron-phonon interaction on the short time-scale, which should properly be done through detailed Green functions simulations. However, this is a difficult problem that has not yet been solved completely. Here we introduce a response function, which is found from more detailed studies, as a descriptor for the finite time taken to complete the electron-phonon interaction. In a sense, then, this approach becomes a first approximation in replacing the Boltzmann equation with a weak scattering formulation of a proper quantum kinetic equation for the density matrix [9]. This response function is then used in the ensemble Monte Carlo process as a secondary rejection process in which scattering events are rejected for short times [27, 28]. Thus, a smooth overall approach to the long-time limit (the Fermi golden rule) is given by the probability function

$$P(t, t_0) = 1 - \exp[-(t - t_0)/\tau]. \quad (7)$$

The response function depends upon an effective collision duration, which has been taken from the Green function calculations of Lipavsky *et al* [29].

In figure 2, the occupations of the X and L valleys in GaAs, after a femtosecond laser pulse, are shown as a function of time after the onset of the laser. The phonon in the Γ -X transfer is the one most heavily affected by the finite collision duration (it is the highest energy phonon). This reduces the number of electrons transferred to the X valleys by almost a factor of 2. However, this also propagates through to the L valley population as well. One important aspect is that this result will certainly affect estimates of intervalley phonon coupling constants that have been made based upon population measurements.

5. Raman scattering in nanostructures

Recently, Grann *et al* [30] have performed a single-particle electronic Raman scattering experiment using a

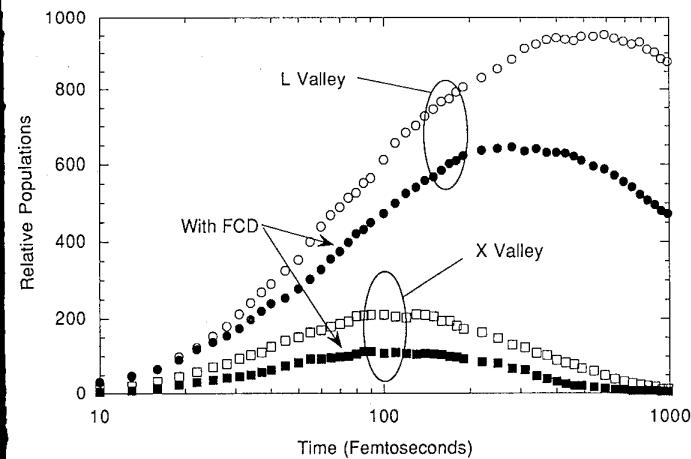


Figure 2. The relative populations of the X and L valleys for a 2.0 eV femtosecond laser excitation, showing the effects of the finite collision duration.

picosecond laser to excite electrons in an AlAs/GaAs/AlAs p-i-n structure. The exciting pulse is also used as a probe for the scattered photons which collide with the photo-excited electrons. The single-particle electronic Raman spectrum thus obtained is due to the energy shift of scattering photons interacting with the electrons, and gives a measure of the velocity distribution of the photoexcited electrons. Such a measurement can be simulated by coupling the ensemble Monte Carlo procedure into a finite structure in real space. The EMC simulation is performed on carriers only in a small sample of the device defined by the volume $W \times W \times L_z$, where W is calculated such that we have only N_{elec} electrons in the simulation. $L_z = 1.0 \mu\text{m}$ is the length of the intrinsic region. Electrons are generated by band-to-band transitions from the heavy-hole, light-hole and split-off valence bands proportional to the respective densities of states of the initial valence state, and the final Γ conduction band state. To account for the device geometry, the z -positions of the simulated electrons are kept track of, and any electron which travels beyond the device dimension in the z -direction is discarded from the simulation. Eventually, all electrons will be emptied out of the simulation volume by the electric fields, and will collect at the contacts at positions $z = 0$ and $1 \mu\text{m}$.

Figure 3 shows the calculated shift of the velocity distribution function, as it would be measured by the Raman signal from the exciting laser pulse (which here is a 3 ps pulse) for electrons in the Γ valley with applied field (the built-in field of the p-i-n structure is 15 kV cm^{-1}). Since the Raman scattered beam is attenuated, the contribution of electrons in the sample decreases with increasing z -position. This reduces the effect of higher-velocity electrons, which have accelerated with the field and travelled further within the sample in the $+z$ direction. Also, the pulse shape causes initial and long-time contributions to be cut off, and the transient information is integrated into the distribution function. The average velocity shows a weak effect of the lower velocity L and X valley electrons. The velocities derived

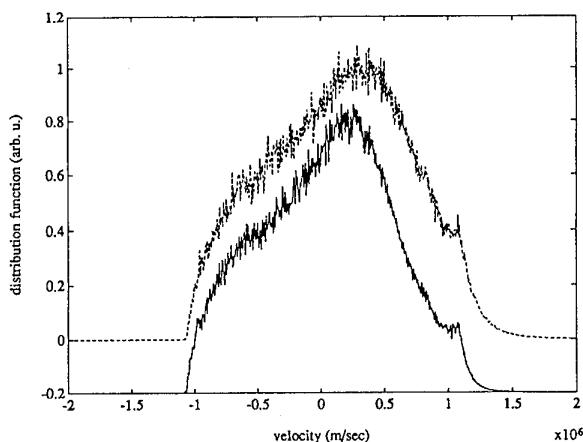


Figure 3. Velocity distribution function from the simulated Raman signal for reverse bias electric fields of 15 kV cm^{-1} (lower, full curve) and 25 kV cm^{-1} (upper, broken curve). The experiment is in a back-scattering configuration.

from the scattered probe pulse spectrum were found to be $4.3 \times 10^4 \text{ cm s}^{-1}$ and $1.32 \times 10^5 \text{ cm s}^{-1}$ for electric fields of 15 and 25 kV cm^{-1} respectively, while distribution functions obtained using only the Γ valley electron velocities show slightly higher values at $4.6 \times 10^4 \text{ cm s}^{-1}$ and $1.36 \times 10^5 \text{ cm s}^{-1}$. Grann *et al* measure the shift of the Raman spectra peak with electric field to define their average velocities and find the values to be approximately $1 \times 10^5 \text{ cm s}^{-1}$ and $2 \times 10^5 \text{ cm s}^{-1}$ respectively for these fields. The corresponding values from figure 3 are found to be somewhat higher at around $2 \times 10^5 \text{ m s}^{-1}$ and $3 \times 10^5 \text{ m s}^{-1}$, but the statistical noise in the simulation makes it difficult to determine the shift of the scattered signal peak precisely [31].

Acknowledgments

This work is the result of contributions from a number of colleagues and former colleagues. The author would especially like to acknowledge the collaborations of J R Barker, S Günçer, H Hida, R P Joshi, M-J Kann, A Kriman, K T Tsien and S Yamaguchi, and the support of the Office of Naval Research.

References

- [1] Shah J, Deveaud B, Damen T C, Tsang W T, Gossard A C and Lugli P 1987 *Phys. Rev. Lett.* **59** 2222
- [2] Hackenberg W and Fasol G 1989 *Solid-State Electron.* **32** 1247
- [3] Fasol G, Hackenberg W, Hughes H P, Ploog K, Bauser E and Kano H 1990 *Phys. Rev. B* **41** 1461
- [4] Ulbrich R G, Kash J A and Tsang J C 1989 *Phys. Rev. Lett.* **62** 949
- [5] Wang W B, Ockman N, Yan M and Alfano R R 1989 *Solid-State Electron.* **32** 1337
- [6] Mickevicius R and Reklaitis A 1990 *Semicond. Sci. Technol.* **5** 805

- [7] Pozhela J and Reklaitis A 1980 *Solid-State Electron.* **23** 927
- [8] Littlejohn M A, Hauser J R and Glisson T H 1977 *J. Appl. Phys.* **48** 4587
- [9] Barker J R and Ferry D K 1979 *Phys. Rev. Lett.* **42** 1779
- [10] Fetter A L and Walecka J D 1971 *Quantum Theory of Many-Particle Systems* (New York: McGraw-Hill)
- [11] Tran Thoai D B and Haug H 1993 *Phys. Rev. B* **47** 3574 and references therein
- [12] See, for example, Kocevar P 1994 *Semicond. Sci. Technol.* **9** 1972
- [13] For a review, see D F Ferry and H L Grubin 1994 *Solid State Physics* ed H Ehrenreich at press
- [14] Kuhn T, Binder E, Rossi F, Lohner A, Rick K, Leisching P, Leitenstorfer A, Elsaesser T and Stolz W 1993 *Coherent Optical Interactions in Semiconductors* at press
- [15] Peyghambarian N, Koch S W, Lindberg M, Fluegel B and Joffre M 1988 *Phys. Rev. Lett.* **62** 1185
- [16] Kuznetsov A V 1991 *Phys. Rev. B* **44** 8721, 13381
- [17] See, for example, the discussion of A P Jauho 1992 *Quantum Transport in Semiconductors* ed D K Ferry and C Jacoboni (New York: Plenum) p 179
- [18] Henning P A 1990 *Nucl. Phys. B* **342** 345
- [19] Hall A G 1975 *J. Phys. A: Math. Gen.* **8** 214
- [20] Tikhodeev S G 1982 *Sov. Phys.-Dokl.* **27** 492, 624
Kukharenko Yu A and Tikhodeev S G 1982
Sov. Phys.-JETP **56** 831
- [21] Lugli P and Ferry D K 1985 *IEEE Electron Device Lett.* **6** 25; 1985 *Appl. Phys. Lett.* **46** 594; 1985 *Physica* **134B** 364
- [22] Jacoboni C 1976 *Proc. 13th Int. Conf. on Physics of Semiconductors* (Rome: Marves) p 1195
- [23] Lugli P and Ferry D K 1986 *Phys. Rev. Lett.* **56** 1295
- [24] Ferry D K, Kann M-J, Kriman A M and Joshi R P 1991 *Comput. Phys. Commun.* **67** 119
- [25] Kriman A M, Kann M J, Ferry D K and Joshi R 1990 *Phys. Rev. Lett.* **65** 1619
- [26] Becker P C, Fragnito H, Brito-Cruz C, Shah J, Fork R L, Cunningham J E, Henry J E and Shank C V 1988 *Phys. Rev. Lett.* **61** 1647
- [27] Ferry D K, Kriman A M, Hida H and Yamaguchi S 1991 *Phys. Rev. Lett.* **67** 633
- [28] Hida H, Yamaguchi S, Kriman A M and Ferry D K 1992 *Semicond. Sci. Technol.* **7** B154
- [29] Lipavsky P, Khan F S, Kalvová A and Wilkins J W 1991 *Phys. Rev. B* **43** 6650
- [30] Grann E, Hsieh S J, Tsen K T, Sankey O F, Maracas G, Salvador A and Morkoc H 1994 *Appl. Phys. Lett.* **64** 1230.
- [31] Günçer S and Ferry D K 1994 *SPIE Proc.* **2142** at press

Long-lived excitonic ground states in GaAs/AlAs coupled quantum well structures

A Zrenner, L V Butov† and M Hagn

†Walter Schottky Institut, TU München, Am Coulombwall, D-85748 Garching, Germany

Abstract. We report optical experiments on electric-field-tunable GaAs/AlAs coupled quantum well structures in the regime of the electric-field-induced Γ -X transition. The formation of real- and k -space indirect excitons causes a huge increase of the exciton lifetime, which is the basis for the phenomena discussed here. We first use the energetically tunable X-point state in the AlAs layer as an internal energy spectrometer to prove the existence of natural quantum dots in the neighbouring GaAs quantum well. Secondly we report on low-temperature cw and time-resolved magneto-optical experiments in the purely indirect regime, designed to search for condensation phenomena of indirect excitons in k -space.

Since the introduction of III-V quantum well (qw) structures most of the research concerning the optical properties of those materials has been concentrated on direct-gap systems such as GaAs/AlGaAs qws. Only in the field of GaAs/AlAs short-period superlattices (SPSL) [1] has the indirect-gap material AlAs attracted appreciable interest. As a function of the individual layer width the lowest energy gap of a GaAs/AlAs layer sequence can be tuned from direct to indirect. Close to the critical layer thickness of the direct–indirect crossover the application of additional electric field [2] or hydrostatic pressure [3] results in a direct–indirect crossover (Γ -X transition in GaAs/AlAs). Recently, GaAs/AlAs coupled qw (cqw) structures have also been introduced [4].

In such structures the excitonic ground state can be tuned from direct to indirect, in both real and k space, by an electric field. The cqws are embedded between two i-Al_{0.48}Ga_{0.52}As layers (see figure 1(a)). Electric tuneability is obtained via additional n⁺ GaAs layers on both sides of the intrinsic region [5]. The structures have been grown by molecular beam epitaxy (MBE). The nominal GaAs qw width is 30 Å. For this width E₀^Γ is almost degenerate with E₀^X. Starting with this alignment the Γ -X transition can be easily obtained by application of an electric field. Three different samples with AlAs layer widths of 30, 40 and 50 Å have been grown. The photoluminescence (PL) energy as a function of V_B for a GaAs/AlAs 30 Å/50 Å structure (full squares) is shown in figure 1(b). The full curves represent theoretical results for excitonic transitions (no Γ -X coupling), assuming the indicated well width [5]. The Γ -X transition occurs at V_B = 0.25 V. One should note the huge enhancement of the exciton lifetime in the indirect regime, which can reach

500 ns as compared with less than 500 ps in the direct regime [4], and also the linear Stark shift which is caused by the electric field acting on the spatially separated electron–hole system.

In the first part of this paper we use our cqw structure as an electric-field-tunable energy spectrometer and charge reservoir to explore the influence of natural interface disorder on the electronic level spectrum. The investigation of the microscopic structure of heterointerfaces has been the subject of many past publications [6–16]. Those contributions can be divided into a group of investigations on optical properties [6–14] and on atomic structure [15, 16]. Whereas by optical techniques the existence of huge monatomically smooth islands with a size of up to several micrometres has been claimed in high-quality growth-interrupted qws [11], structural methods have revealed substantial roughness and alloy fluctuations on an atomic length scale [15, 16]. Recently, optical data have also been shown to be inconsistent with the existence of huge monatomically smooth islands [14]. Currently, interface roughness with an amplitude of at least one monolayer (ML) in the growth direction is believed to appear on a broad range of length scales. In a narrow qw well width fluctuations of several monolayers will result in sizable lateral potential variations. On this basis it is justified to describe a narrow qw as a disordered array of quantum dots with arbitrary dimensions.

To demonstrate the existence of zero-dimensional states, we make use of electric-field-tunable AlAs/GaAs coupled qw structures to overcome the intrinsic lifetime limitations of single qws and to benefit from resonant carrier injection into GaAs quantum dot levels. From model calculations we know that a width variation between 10 and 12 ML will result in a 43 meV shift of the

† Permanent address: Institute of Solid State Physics, Russian Academy of Science, 142432 Chernogolovka, Moscow district, Russia.

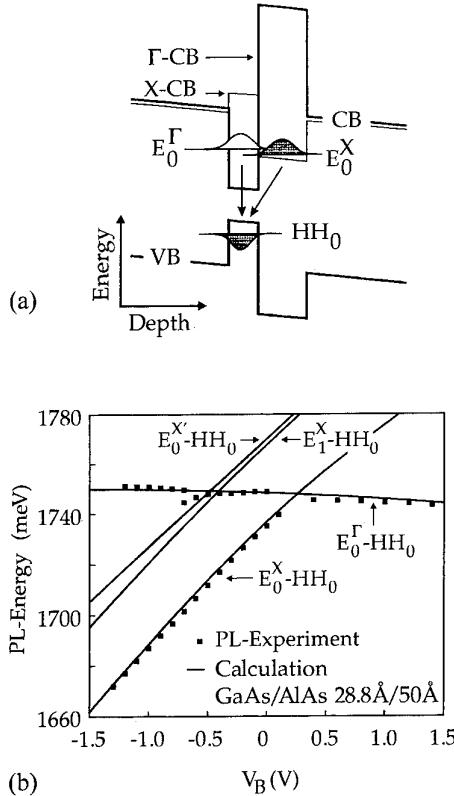


Figure 1. (a) Band diagram of a real- and k -space indirect AlAs/GaAs quantum well for negative V_g . (b) V_g dependence of the direct and indirect transitions.

direct transition ($E_0^\Gamma - HH_0$) in the GaAs quantum well. The major part of this (36 meV) appears in E_0^Γ . Since E_0^X can be tuned by more than 70 meV with respect to E_0^Γ , the X-point level E_0^X can be used as an internal energy spectrometer to map out the local energy minima of the GaAs quantum well.

In figures 2(a) and (b) we show PL results in the indirect regime ($V_B = -0.2$ V) for two different sampling areas. For the spectrum shown in figure 2(a) the laser was focused to a diameter d_L of 100 μm. The direct ($E_0^\Gamma - HH_0$) and the indirect ($E_0^X - HH_0$) transitions are observed. Whereas the direct line has a Gaussian shape as expected for the inhomogeneously broadened line of a quantum well grown without interruptions, a structured tail is detected on the low-energy side of the indirect line. The structure in this tail appears tremendously enhanced as the laser is focused down to $d_L = 2$ μm and moved to an appropriate position in the x - y plane of the mesa diode (see figure 2(b)). New extremely narrow emission lines (labelled from $n = 1$ to 7) emerge from the background of the indirect line. The FWHM of those new emissions is about 0.5 meV in the observed spectrum, limited by the choice of slits in the spectrometer (the bare linewidth is about 0.2 meV). Both the narrow linewidth and the level sequence are reminiscent of an emission spectrum from a fully quantized system such as a single quantum dot [17]. The low-energy cut-off of the narrow emission lines is about 30 meV below the position of the direct PL and 38 meV below the direct absorption edge, which cor-

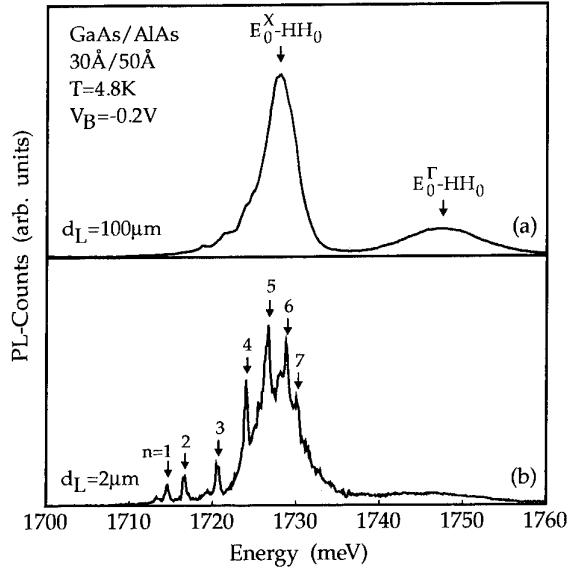


Figure 2. (a) PL response of a GaAs/AlAs 30 Å/50 Å coupled quantum well structure in the indirect regime for $V_B = -0.2$ V. The diameter of the optically probed area d_L is 100 μm. (b) Same as (a) only with $d_L = 2$ μm. New narrow lines (labelled from $n = 1$ to 7) appear in the region of the indirect PL.

responds to an almost 2 ML variation in well width. It can be further shown that a 1000 Å wide and 20 meV deep Gaussian-shaped potential has an almost similar level spectrum as the one shown in figure 2(b). The observed level spectrum changes as we choose a different position in the x - y plane of the mesa diode.

Results from a GaAs/AlAs 30 Å/40 Å structure as a function of V_B are shown in figure 3. Compared with figure 2(b) the observed level spectrum is more complicated. The total number of observed lines decreases with decreasing V_B . The linear Stark shift of the indirect line is still evident as a global red-shift of the background from which the narrow emission lines emerge. There is, however, only negligible Stark shift on the position of the narrow emission lines. From this, and also from the strength of the emission lines, we conclude that the origin of the new lines cannot be from an indirect recombination process. We rather think that we observe direct transitions from zero-dimensional states in local potential minima of the GaAs quantum well. In general we expect to find more than one local potential minimum in an optically probed area of $d_L = 2$ μm. For small Γ - X separations carrier injection into a big variety of shallow and deep local potential minima is possible and a complex superposition of sets of narrow emission lines results ($V_B = 0$ V). For large separations ($V_B = -0.6$ V) only sufficiently deep potential minima can be populated and the spectra simplify. As sketched in the inset of figure 3, electron-hole pairs are photoexcited selectively in the GaAs quantum well. The electrons transfer subsequently in a fast Γ - X relaxation process into the electric-field-tunable, long-lived E_0^X state of the AlAs quantum well. In this sense the X-point level E_0^X is a tuneable charge reservoir which can be used to inject

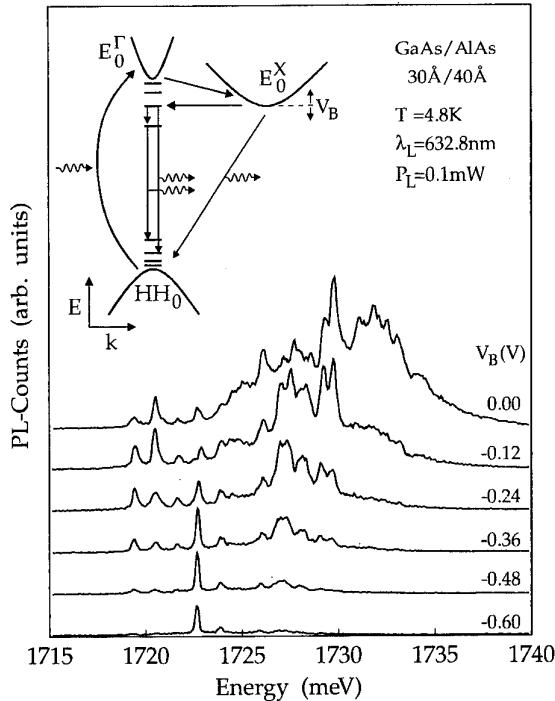


Figure 3. PL response in the indirect regime as a function of V_B . The diameter of the optically probed area d_L is 2 μm . The relevant relaxation and recombination processes after optical excitation in the GaAs qw ($\text{HH}_0 \rightarrow \text{E}_0^\Gamma$) are shown in the inset.

electrons resonantly back from the X point into local energy minima of the direct GaAs qw at the Γ point.

Combining the principle of resonant charge injection with spatially resolved measurements we are able to map out the in-plane potential fluctuations. As indicated in the inset of figure 4 we have performed a series of spatially resolved PL measurements over an area of 45 $\mu\text{m} \times 30 \mu\text{m}$ with a pitch of 3 μm in x and y directions and a laser spot diameter of about 2 μm . The position and strength of the narrow emission lines in figure 4(a) ($V_B = 0.1 \text{ V}$) is a strong function of the in-plane coordinates. Narrow emission lines with different amplitudes can be detected over almost the entire scanned area. There also exist larger clusters with different amounts of structure in the PL spectrum. All features are strictly reproducible. The data are consistent with our earlier proposed picture of the GaAs qw as a disordered array of arbitrarily sized quantum dots. The amplitude of the well width fluctuations can be explored by varying the injection energy via V_B . Results for $V_B = -0.1 \text{ V}$ are shown in figure 4(b). As compared with $V_B = 0.1 \text{ V}$ the injection energy, which is given by the energetic position of the indirect line, is lowered by about 6 meV. At this reduced injection energy allowed energy states in the GaAs qw are harder to find. Over most of the scanned area only the undistorted indirect PL line is observed. Narrow emission lines emerge only at some selected locations. Those are the regions where the local width of the GaAs qw has maxima.

Since our dots are in principle very asymmetric with

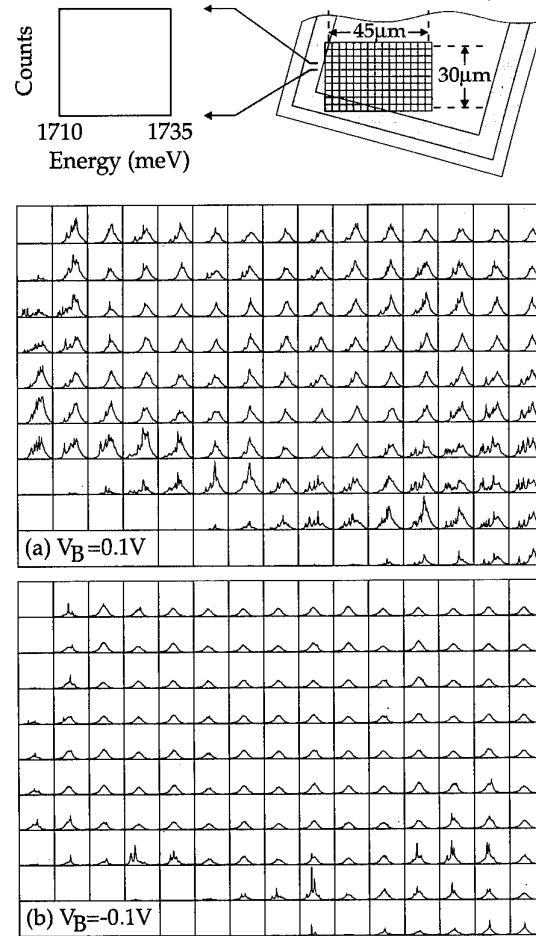


Figure 4. Topology of the PL response for two different V_B . An area of 45 $\mu\text{m} \times 30 \mu\text{m}$ has been scanned with a pitch of 3 μm in x and y. A PL spectrum is plotted for each position. (a) $V_B = 0.1 \text{ V}$: huge variations in the PL response, indicating substantial potential fluctuations below the injection energy. (b) $V_B = -0.1 \text{ V}$: only a few very deep potential minima fall below the reduced injection energy.

strong confinement in the z direction (30 Å) and weak confinement in the x-y plane, their response in a parallel ($B_{||}$) or perpendicular (B_{\perp}) magnetic field has to be totally different. Experimentally we performed our magneto-optical measurements in a superconducting magnet using an optical fibre. Since the optically probed area was about 10 μm in diameter, emission lines from several local potential minima are contained in the spectra. In our experiments (up to 8 T for $B_{||}$) we find negligible influence of $B_{||}$ on the position of the narrow lines ($\Delta E < 0.1 \text{ meV}$). According to $\Delta E = e^2 \langle z_i^2 \rangle B_{||}^2 / 2m^*$ we calculate for E_0^Γ and HH_0 at $B_{||} = 8 \text{ T}$ a total diamagnetic shift of only 0.2 meV (e is the electron charge, $\langle z_i^2 \rangle$ the expectation value of z_i^2 , i the subband index and m^* the effective mass).

For B_{\perp} the positions of the observed emission lines are plotted in figure 5. Complicated level shifts, splittings and anticrossings are observed. The complexity of the data is partly due to the fact that contributions from several potential minima are contained. If we concentrate, however, on the three to four emission lines on the

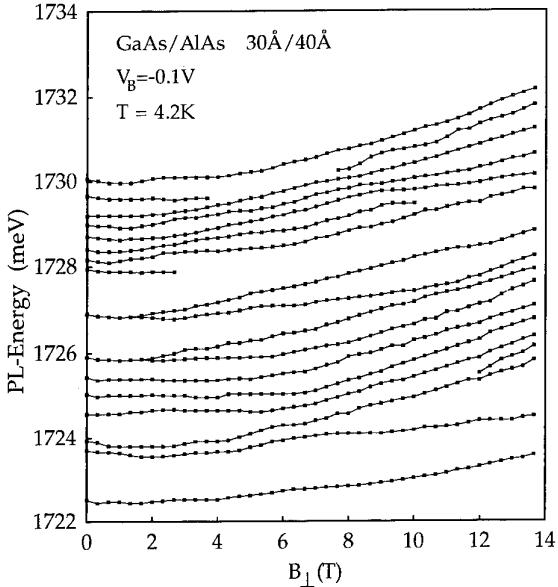


Figure 5. Position of the narrow emission lines versus B_{\perp} .

low-energy end of the spectrum, which we think originate from a single potential minimum, we find qualitatively very similar behaviour to that theoretically predicted by Halonen *et al* [18] for excitons in quantum dots in magnetic fields. For the present work we can therefore conclude that we do indeed observe quantum dots with the proposed geometry, namely weak confinement in the x - y plane and strong confinement in the z direction.

In the second part of the paper we present low-temperature magneto-optical experiments designed to look for the ground state of a cold, dilute 2D exciton system. For those investigations we operate our CQWS again in the indirect regime to benefit from the long exciton lifetimes. We concentrate now, however, on the properties of the indirect excitons. Because of the huge electron mass of the associated AlAs X-point state the previously described well width fluctuations result in considerably smaller potential variations for indirect excitons as compared with the lighter direct excitons.

The electron-hole (e-h) interaction in the neutral e-h system has been predicted to cause the condensation of bound e-h pairs (excitons) into an excitonic insulator state [19]. In the case of a dilute excitonic gas the excitons can be considered as rigid bosons and the condensation is analogous to Bose-Einstein condensation, while in the case of the dense e-h system the condensed state is analogous to the BCS superconducting state. The transition between those two limits is smooth [20]. The condensation condition can be achieved only if the temperature of the excitons is below a critical temperature T_c , a criterion which is hard to achieve experimentally.

Experimental work on exciton condensation in bulk semiconductors was concentrated mainly on the analysis of the photoluminescence (PL) lineshape and the transport properties of excitons [21-24]. In 2D systems the precursor of exciton condensation, namely the formation of excitons in a dense e-h magnetoplasma, has been

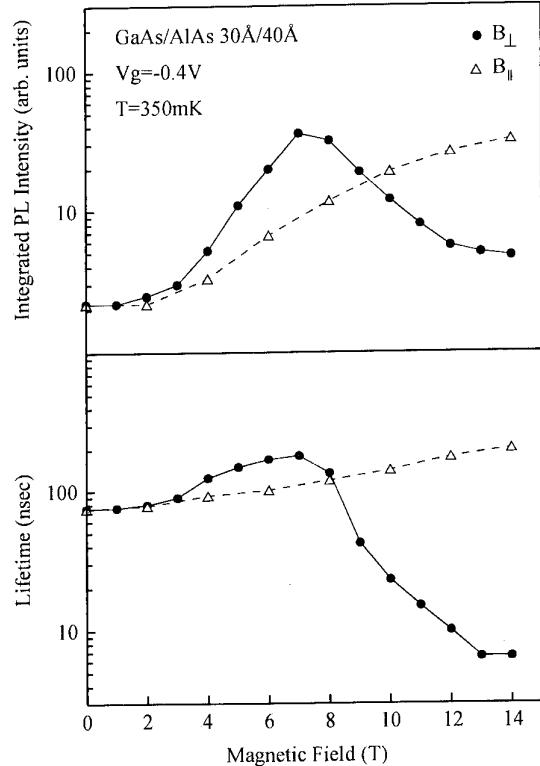


Figure 6. Integrated PL intensity and decay time in the indirect regime ($V_g = -0.4$ V) at $T = 350$ mK versus B_{\perp} (full circles) and B_{\parallel} (open triangles) in the logarithmic scale.

observed in InGaAs QWS [25]. Due to the high rate of recombination, the temperature of the photoexcited carriers, however, is higher than T_c . A strong suppression of the recombination rate can be achieved in indirect CQWS [4]. Previously reported variations of the PL linewidth [26], however, cannot be interpreted in terms of exciton condensation since the linewidth was found to be dominated by random potential fluctuations [27].

In electric-field-tunable AlAs/GaAs CQW structures the lifetime is sufficiently long to allow for a thermalization of the indirect excitons down to temperatures below 1 K. A perpendicular magnetic field increases the exciton binding energy and suppresses the kinetic energy of the excitons, which leads to an increase of T_c [28, 29]. For spatially separated electrons and holes the condensed excitonic phase has been predicted to be the ground state for $d < l_H$ (d is an interlayer separation, l_H is a magnetic length) [30] at $T < T_c$.

In the indirect regime the integrated PL intensity (I_{PL}) drops almost two orders of magnitude compared with the direct regime. In the indirect regime the radiative lifetime (τ_r) is much larger than the non-radiative lifetime (τ_{nr}) and the total recombination lifetime (τ) is determined by τ_{nr} ($\tau \approx \tau_{nr}$). In time-resolved measurements we have determined τ to be 100 ns for $V_g = -0.5$ V, for example.

The magnetic field causes strong changes of both I_{PL} and τ in the indirect regime. This is shown in figure 6 for B_{\perp} and B_{\parallel} . With increasing B_{\perp} both I_{PL} and τ first

increase and then decrease dramatically. B_{\parallel} results in a monotonic increase of I_{PL} and τ . These effects disappear at temperatures above 5 K where only a small monotonic increase of I_{PL} (about a factor of 2 for $B_{\perp} = 14$ T) is observed. The variations of I_{PL} and τ are correlated, which means that τ_{nr} is subjected to a strong change by the magnetic field. The non-radiative recombination occurs at non-radiative recombination centres (NC). The low-temperature diffusivity of excitons in QWS is determined by interface roughness scattering and is very low [31], the rate of non-radiative recombination at the NC is determined by transport to the NC [32]. An increase of B_{\perp} first leads to a reduction of the exciton diffusivity and then to a strong increase. Increasing B_{\parallel} leads to a monotonic reduction of the diffusivity.

The observed reduction of the exciton diffusivity in a magnetic field can be qualitatively explained by the coupling between the magnetoexciton motion and its internal structure. An exciton moving in a perpendicular field acquires a dipole moment because the Lorentz force acts on electrons and holes in opposite directions. This leads to increased exciton scattering by interface roughness and thus to the observed reduction of the exciton diffusivity. This mechanism, however, cannot explain the dramatic and sudden increase of the exciton diffusivity at high B_{\perp} . This striking anomaly in the magnetoexciton transport seems to be connected with coherent effects in the exciton system. This is confirmed by the fact that the effect disappears at higher temperatures ($T > 5$ K). For weakly interacting bosons a transition to superfluid flow is expected for $T < T_c$.

Another spectacular effect is observed in the indirect regime at finite B_{\perp} , namely a huge broad band noise in I_{PL} . The time dependence of the PL signal at $B_{\perp} = 11$ T and $V_g = -0.5$ V is shown in figure 7. For comparison data with about the same PL intensity for $B = 0$ T are also included (achieved at six times higher W_{ex}). The noise amplitude at $B_{\perp} = 11$ T is anomalously large as compared with the case of $B = 0$ T, which represents an upper limit for noise from Poisson statistics for photons from random single-exciton decay. The power spectrum of the anomalously large noise has a broad band spectrum (see inset of figure 7).

The appearance of the huge noise is strong evidence for the presence of coherence in the exciton system. The noise amplitude is known to be inversely proportional to the number of statistically independent entities in a system [33]. We believe that those macroscopic entities in the exciton system are given by domains of condensed indirect excitons. The PL signal of condensed excitons is much higher than that of uncondensed ones [34]. This is mainly due to the fact that the oscillator strength of the excitonic PL is increased with an increase of the coherent area [35, 36]. The formation and disappearance of condensed domains results, therefore, in changes of I_{PL} . Noise is observed only at low temperatures. The characteristic temperature dependence of the noise amplitude $\langle \delta I_{PL} \rangle / \langle I_{PL} \rangle$ is shown in the inset of figure 7. The decrease of the noise amplitude with increasing temperature reflects a reduction of the average size of

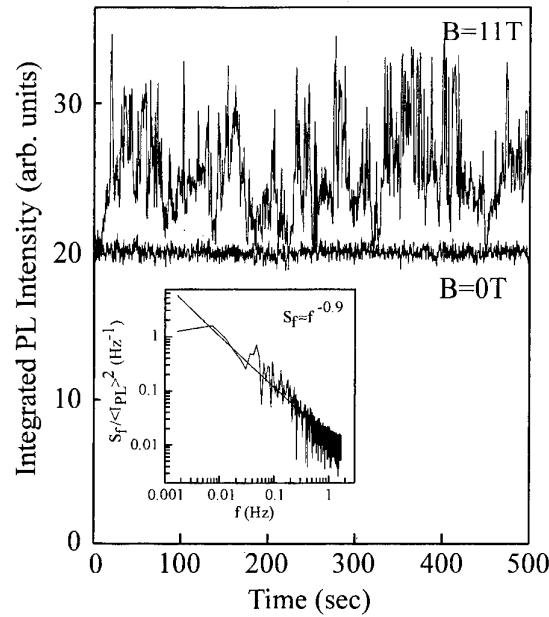


Figure 7. The time dependence of the PL signal in the indirect regime ($V_g = -0.5$ V) at $T = 350$ mK for $B_{\perp} = 11$ T and at $B = 0$. The excitation density for $B = 0$ is six times higher than for $B_{\perp} = 11$ T to get approximately the same signal level. Inset: power spectrum of the noise for $B_{\perp} = 11$ T.

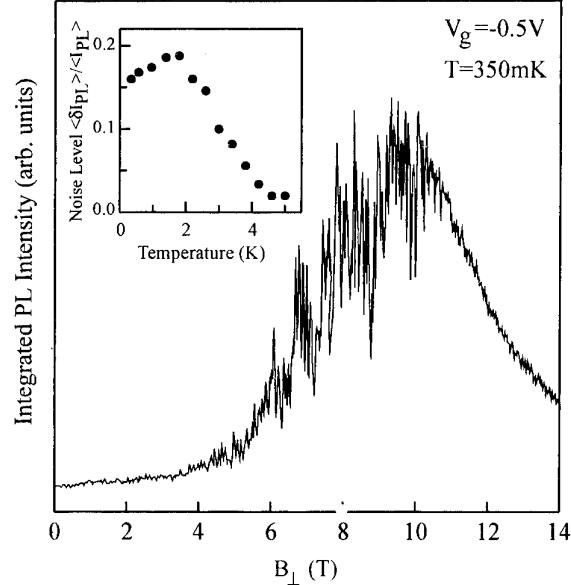


Figure 8. PL intensity in the indirect regime ($V_g = -0.5$ V) at $T = 350$ mK versus B_{\perp} . Inset: corresponding temperature dependence of the noise level $\langle \delta I_{PL} \rangle / \langle I_{PL} \rangle$ at $B = 9$ T.

condensed domains. Noise is observed only in the excitonic regime and disappears at high carrier densities. The range in B_{\perp} over which noise is observed is shown in figure 8. Noise is strongly suppressed at the highest fields corresponding to the reduced excitonic lifetime (see figure 6). In this range of B_{\perp} the diffusivity of excitons is

high. Therefore we suppose that even if condensed domains are formed they exist only for a short time due to the fast transport of excitons to the NCS.

In conclusion we have shown that long-lived excitons in GaAs/AlAs cQWs can be used for a variety of new and interesting experiments. In the first part we used the energetically tuneable X-point state in the AlAs layer as an internal energy spectrometer to prove the existence of natural quantum dots in the neighbouring GaAs quantum well. In the second part we have provided strong evidence for a Bose-Einstein condensed excitonic ground state at low temperatures.

Acknowledgments

We want to acknowledge the participation of G Abstreiter, G Böhm, G Tränkle and G Weimann in this work. We would like to thank G E W Bauer, A B Dzyubenko, Al L Efros, V D Kulakovskii and A Odintsov for many helpful discussions. This work was supported in part by the BMFT (Photonik project number: 01BV219) and by the DFG (SFB 348). LVB thanks the FVS foundation for financial support.

References

- [1] Finkman E, Sturge M D and Tamargo M C 1986 *Appl. Phys. Lett.* **49** 1299
- [2] Meynadier M H, Nahory R E, Worlock J M, Tamargo M C, Miguel J L and Sturge M D 1988 *Phys. Rev. Lett.* **60** 1338
- [3] Skolnick M S, Smith G M, Spain I L, Whitehouse R C, Herbert D C, Whittaker D M and Reed L J 1989 *Phys. Rev. B* **39** 11191
- [4] Zrenner A, Leeb P, Schäfer J, Böhm G, Weimann G, Worlock J M, Florez L T and Harbison J P 1992 *Surf. Sci.* **263** 496
- [5] Zrenner A 1992 *Festkörperprobleme/Advances in Solid State Physics* vol 32 ed U Rössler (Braunschweig: Vieweg) pp 61–80
- [6] Weisbuch C, Dingle R, Gossard A C and Wiegmann W 1981 *Solid State Commun.* **38** 709
- [7] Goldstein L, Horikoshi Y, Tarucha S and Okamoto H 1983 *Japan. J. Appl. Phys.* **22** 1489
- [8] Reynolds D C, Bajaj K K, Litton C W, Yu P W, Singh J, Masselink W T, Fischer R and Morkoc H 1985 *Appl. Phys. Lett.* **46** 51
- [9] Miller R C, Tu C W, Sputz S K and Kopf R F 1986 *Appl. Phys. Lett.* **49** 1245
- [10] Voillot F, Madhukar A, Kim J Y, Chen P, Cho N M, Tang W C and Newman P G 1986 *Appl. Phys. Lett.* **48** 1009
- [11] Bimberg D, Christen J, Fukunaga T, Nakashima H, Mars D E and Miller J N 1987 *J. Vac. Sci. Technol. B* **5** 1191
- [12] Petroff P M, Cibert J, Gossard A C, Dolan G J and Tu C W 1987 *J. Vac. Sci. Technol. B* **5** 1204
- [13] Kohl M, Heitmann D, Tarucha S, Leo K and Ploog K 1989 *Phys. Rev. B* **39** 7736
- [14] Warwick C A, Jan W Y, Ourmazd A and Harris T D 1990 *Appl. Phys. Lett.* **56** 2666
- [15] Ourmazd A, Taylor D W, Cunningham J and Tu C W 1989 *Phys. Rev. Lett.* **62** 933
- [16] Salemink H W M and Albrektsen O 1993 *Phys. Rev. B* **47** 16044
- [17] Brunner K, Bockelmann U, Abstreiter G, Walther M, Böhm G, Tränkle G and Weimann G 1992 *Phys. Rev. Lett.* **69** 3216
- [18] Halonen V, Chakraborty T and Pietiläinen P 1992 *Phys. Rev. B* **45** 5980
- [19] Keldysh L V and Kopaev Yu E 1965 *Sov. Phys.—Solid State* **6** 2219
- [20] Comte C and Nozieres P 1982 *J. Physique* **43** 1069; Nozieres P and Comte C 1982 *J. Physique* **43** 1083
- [21] Mysyrowicz A and Benoit à la Guillaume C 1980 *Phys. Rev. Lett.* **45** 1970
- [22] Timofeev V B, Kulakovskii V D and Kukushkin I V 1983 *Physica* **117 & 118** 327
- [23] Peyghambarian N, Chase L L and Mysyrowicz A 1983 *Phys. Rev. B* **27** 2325
- [24] Snoke D W, Wolfe J P and Mysyrowicz A 1990 *Phys. Rev. Lett.* **64** 2543; Fortin E, Fafard S and Mysyrowicz A 1993 *Phys. Rev. Lett.* **70** 3951
- [25] Butov L V and Kulakovskii V D 1991 *JETP Lett.* **53** 466
- [26] Fukuzawa T, Mendez E E and Hong J M 1990 *Phys. Rev. Lett.* **64** 3066
- [27] Kash J A, Zachau M, Mendez E E, Hong J M and Fukuzawa T 1991 *Phys. Rev. Lett.* **66** 2247
- [28] Lerner V and Lozovik Yu E 1981 *Sov. Phys.—JETP* **53** 763
- [29] Bauer G E W 1992 ‘Optics of Excitons in Confined Systems’ (*Inst. Phys. Conf. Ser.* 123) ed A D’Andrea *et al* (Bristol: Institute of Physics) p 283
- [30] Yoshioka D and MacDonald A H 1990 *J. Phys. Soc. Japan* **59** 4211
- [31] Hilmer H, Forchel A, Sauer R and Tu C W 1990 *Phys. Rev. B* **42** 3220
- [32] Abakumov V N, Perel V I and Yassievich I N 1991 *Nonradiative Recombination in Semiconductors* ed V M Agranovich and A A Maradudin (Amsterdam: North-Holland)
- [33] Altshuler B L, Lee P A and Webb R A 1991 *Mesoscopic Phenomena in Solids* ed V M Agranovich and A A Maradudin (Amsterdam: North-Holland)
- [34] Bauer G E W 1992 *Phys. Scri.* **T45** 154
- [35] Rashba E I and Gurgenishvili G E 1962 *Sov. Phys.—Solid State* **4** 759
- [36] Feldmann J, Peter G, Göbel E O, Dawson P, Moore K, Foxon C and Elliott R J 1987 *Phys. Rev. Lett.* **59** 2337

Bloch and localized electrons in semiconductor superlattices

M Helm[†], W Hilber[†], T Fromherz[†], F M Peeters[‡], K Alavi[§] and
R N Pathak[§]

[†] Institut für Halbleiterphysik, Universität Linz, A-4040 Linz, Austria

[‡] Department of Physics, Universiteit Antwerpen, B-2610 Antwerpen, Belgium

[§] Department of Electrical Engineering, Center for Advanced Electron Devices and Systems, University of Texas at Arlington, Arlington, TX 76019, USA

Abstract. The energy spectrum of electrons in semiconductor superlattices consists of Bloch-like miniband states and localized impurity states. Optical transitions between these states are investigated by infrared spectroscopy. The inter-miniband absorption spectrum is dominated by the van Hove singularities of the quasi one-dimensional joint density of states. The detailed shape of the absorption can be understood in terms of the relevant transition matrix elements and the electron concentration, i.e. it depends on whether the first miniband is full or half-filled. We show that this has implications for intra-miniband absorption and the optical sum rules. In addition, a transition between the states localized at the lower edge of the minibands occurs, which can be traced back to the 1s–2p_z hydrogenic donor transition in the low-doping limit. Finally, we discuss the relevance of our observations for the study of the metal–insulator transition in superlattices.

1. Introduction

Semiconductor superlattices represent a fascinating model system for the study of many key issues of solid state physics. The artificial periodicity allows one to probe the electron states and dynamics in a parameter range not accessible in ‘natural’ crystals. The most prominent achievements in the recent past have been the observation of Esaki–Tsu-type [1] negative differential resistance [2–5], which occurs when electrons are accelerated into the negative-mass region of the miniband dispersion, and, related to this, the observation of Bloch oscillations [6] and the Wannier–Stark ladder [7], both of which are discussed elsewhere in this issue. All of these effects represent a different aspect of the same physical phenomenon, the electric-field-induced localization of the electron wavefunction. In order to observe these effects, the superlattice has to be of very high quality, so that the wavefunction at zero electric field is truly extended and coherent over several superlattice periods. This is equivalent to the requirement that the intrinsic width of the minibands is larger than the energy scale of any broadening mechanism, like scattering or potential fluctuations. In this case the electrons exhibit a well developed dispersion along the growth axis (Bloch electrons).

On the other hand, it is well known that if any weak disorder is present each energy band has a set of localized states at its lower edge. In bulk semiconductors with low doping concentration these are just the discrete, hydrogenic impurity states. For higher doping impurity bands are formed, which finally merge with the conduction band

into a tail of localized states. In the case of superlattices there are two basic differences:

(i) The impurity binding energy depends on the position of the impurity along the growth axis [8, 9]. Therefore, if the superlattice is homogeneously doped, an impurity band is formed even for low doping.

(ii) Even the most perfect *real* superlattices exhibit fluctuations of the layer thicknesses and potential heights. This is an additional source of Anderson localization not present in bulk crystalline semiconductors.

In this work we present a study of the optical properties of conduction electrons in n-doped GaAs/AlGaAs superlattices. In contrast to other work, where valence-to-conduction band transitions were investigated [10], we look at optical transitions between minibands within the conduction band (which occur in the mid-infrared spectral range). Both Bloch-like miniband states and localized states can be identified. The miniband states are characterized by their energy dispersion along the superlattice axis. The van Hove singularities in the joint density of states (JDOS) at the centre and the edge of the superlattice Brillouin zone are spectroscopically evidenced through absorption maxima [11]. The relative sizes of these maxima, i.e. the asymmetry, can be understood by considering the proper transition matrix elements. We argue that there is a fundamental connection between the asymmetry and the bandwidth, which is related to the oscillator sum rule. In addition, a transition between localized states, which are located just below the first and second minibands, is observed. We show that it can be

traced back to the transition between the 1s and $2p_z$ hydrogenic donor levels [12, 13].

2. Experiment

Two GaAs/Al_{0.3}Ga_{0.7}As superlattices with different doping concentrations were studied. They were grown by molecular beam epitaxy on semi-insulating GaAs (100) substrates and a 2000 Å wide, undoped GaAs buffer. In both cases the nominal width of the GaAs quantum wells was 75 Å and the width of the AlGaAs barriers 25 Å, corresponding to a superlattice period of $d = 100$ Å. Superlattice no 1 was homogeneously doped n-type, to give an electron concentration of 6×10^{17} cm⁻³, with a total thickness of 2 µm (200 periods), superlattice no 2 was doped to $n = 6 \times 10^{16}$ cm⁻³ and had a thickness of 5 µm (500 periods). Due to the homogeneous doping, band bending can be assumed to be rather small. Sample characterization was performed by x-ray diffraction and Hall measurements, yielding a period of 96.5 Å for both superlattices and a peak mobility of 3000 cm² V⁻¹ s⁻¹ (at 200 K) for sample no 1 and 6000 cm² V⁻¹ s⁻¹ (at 100 K) for no 2.

Infrared absorption measurements were performed with a Bruker IFS 113v rapid-scan Fourier transform spectrometer. The samples were mounted in a liquid-helium flow cryostat, where the sample temperature could be varied between 5 K and 300 K. In order to achieve an active polarization for the intersubband absorption (electric field perpendicular to the layers) the samples were prepared in a multipass waveguide geometry [14]. Then the ratio of active and inactive polarization was measured for the superlattice and a reference substrate in the same geometry. In order to obtain data that were completely free of any system artefacts, a metallic gate (Ti/Au) was deposited on the waveguide samples, and was used to modulate the electron density in the top section of the sample by applying a proper reverse bias.

3. Theory

Figure 1 shows the band structure for the nominal parameters of the superlattices, calculated within the one-band effective-mass approximation. Also shown are the energetic positions of the 1s and $2p_z$ donor states, calculated for independent impurities with the variational method described in [13]. Although the binding energy of impurities in superlattices depends on the position of the impurity atom [8, 9], this variation is smaller than 3 meV in the present case and can therefore be neglected compared with the other energies involved. Moreover, since the doping level is well above the metal-insulator transition (which occurs at $n = 1.5 \times 10^{16}$ cm⁻³ in bulk GaAs), the donors will form impurity bands rather than isolated energy levels. Still, the behaviour of the pure hydrogen-like states allow for some physical insight. In figure 2 the probability densities along the growth axis of the 1s and $2p_z$ states are shown together with the

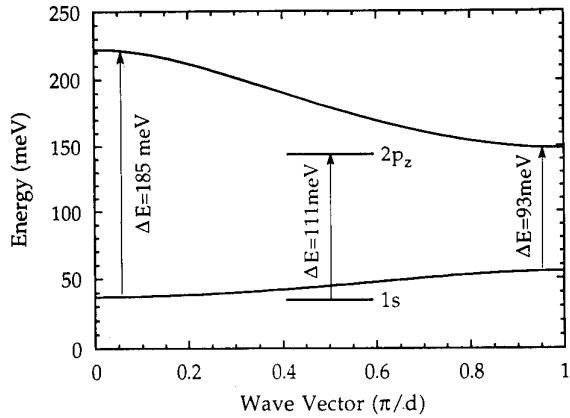


Figure 1. The calculated dispersion for the two lowest minibands. The 1s and $2p_z$ states are included schematically. The transition energies at the centre and the edge of the mini-Brillouin zone as well as the impurity transition energy are indicated.

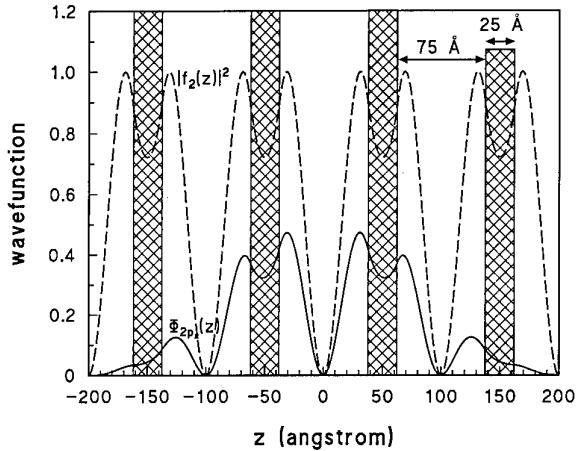
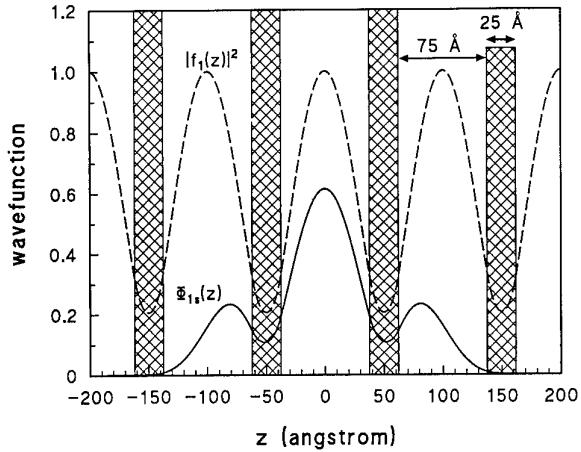


Figure 2. The probability distributions of the 1s and $2p_z$ states (upper and lower panel respectively, full curves) are shown together with the probability distributions of the Bloch electrons at the bottom of the first (upper panel) and second (lower panel) minibands (broken curves). The relative amplitudes of the miniband and impurity states is arbitrary, since the extended states are not normalizable. The barrier regions are indicated by the hatched regions.

probability densities of the Bloch states at the bottom of the first and second minibands (which are the states $|n = 1, k_z = 0\rangle$ and $|n = 2, k_z = \pi/d\rangle$ respectively). It is evident that the donor wavefunctions decay exponentially over a few superlattice periods, but are modulated by the characteristics of their associated Bloch states. In fact, even if a different model is used to describe the localized states, such as a superlattice containing one slightly wider quantum well, the resulting wavefunctions (in the z -direction) turn out to be nearly identical [15].

In order to obtain the absorption coefficient, one has to sum over all transitions from occupied to empty states (extended and localized), taking into account the proper occupation statistics and matrix elements. The problem, however, is that the present single-particle approach is not able to properly account for the nature of the impurity band and the position of the Fermi energy in the case of a degenerately doped semiconductor [16].

Therefore, we will attempt to analyse the absorption spectrum by separating miniband and impurity contributions. Within a single-particle approximation, the absorption coefficient for transitions between the two lowest minibands (extended states) can be expressed as

$$\alpha = \frac{e^2 k T}{\epsilon_0 c \eta \hbar^2 \pi m^* \omega} \int_0^{\pi/d} dk_z |\langle 1 | p_z | 2 \rangle|^2 \times \ln \left(\frac{1 + \exp[(E_F - E_1(k_z))/kT]}{1 + \exp[(E_F - E_2(k_z))/kT]} \right) \times \left(\frac{\Gamma/\pi}{(E_2(k_z) - E_1(k_z) - \hbar\omega)^2 + \Gamma^2} \right). \quad (1)$$

Here k is Boltzmann's constant and η is the refractive index. The other quantities have their usual meaning. The dispersions of the first and second minibands, $E_{1(2)}(k_z)$, as well as the momentum matrix elements (also k_z dependent) are calculated within the above model. Note that for infinite, periodic systems, the $\mathbf{A} \cdot \mathbf{p}$ interaction has to be used for the electron–radiation coupling, whereas the $e\mathbf{E} \cdot \mathbf{r}$ interaction leads to wrong results, if the wavefunction does not vanish in the barriers [17]. The Fermi energy, E_F , has to be determined from the electron concentration for each temperature. It is assumed that each transition at wavevector k_z is broadened in a Lorentzian fashion with the broadening parameter Γ . There are three important contributions determining the absorption coefficient: the joint density of states, the Fermi–Dirac distribution (already integrated over k_x and k_y) and the square momentum matrix element. The JDOS for the present superlattices is shown in figure 3 (dotted curve). The slight asymmetry stems from the fact that the curvature of Kronig–Penney-like bands is different at $k_z = 0$ and $k_z = \pi/d$. For tight-binding bands the shape would be exactly symmetric [13, 18]. In a parabolic model, the JDOS is strictly one dimensional and has $1/\sqrt{E}$ -like singularities. This is in contrast to interband transitions [10], where due to the opposite curvature of electron and hole bands the JDOS exhibits three-dimensional critical points of type M0 and M1 at the minizone centre and edge respectively.

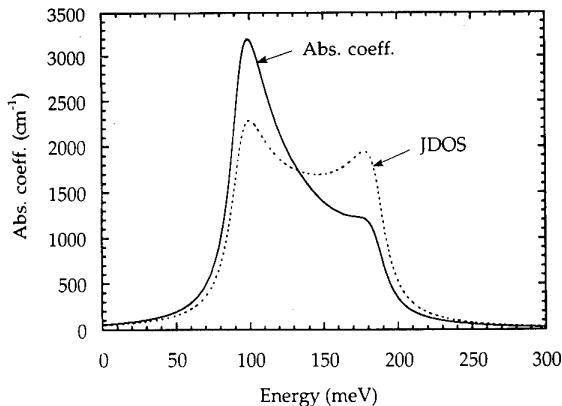


Figure 3. One-dimensional joint density of states (dotted curve) and absorption coefficient according to equation (1) for $\Gamma = 10$ meV, $n = 6 \times 10^{17}$ cm $^{-3}$ and $T = 5$ K (full curve) for the minibands of figure 1. The units of the JDOS are arbitrary.

Rigorously, however, when the non-parabolic increase with energy of the in-plane mass is taken into account, the JDOS for inter-miniband absorption is also three-dimensional with critical points of type M3 (zone centre) and M2 (zone edge).

The second contribution to α is the thermal occupation factor, which becomes particularly important when the first miniband is only partially filled at low temperatures (i.e. E_F lies within the first miniband). This is the case for sample no 2, where E_F is calculated to be 43 meV (neglecting donor occupation effects) at low temperature, whereas for sample no 1 $E_F = 67$ meV. In this sample the first miniband is completely filled and E_F lies in the minigap (which is, of course, no real gap, since it contains states with finite in-plane wavevectors).

The third significant contribution to α is the oscillator strength, f_{12} , which is related to the momentum matrix element by $f_{12} = (2/m^* \hbar \omega_{21}) |\langle 1 | p_z | 2 \rangle|^2$. For the present system we calculate $f_{12} = 0.3$ at the zone centre ($k_z = 0$) and $f_{12} = 2.3$ at the zone edge ($k_z = \pi/d$), which is a variation by a factor of 7.5 (the physical reason for this will be discussed later). This behaviour strongly enhances the low-frequency peak. The resulting absorption coefficient (for sample no 1 at $T = 5$ K) is also shown in figure 3 (full curve).

In a very simplistic picture, the $1s-2p_z$ impurity absorption is modelled as a transition between discrete states, which is strictly valid only for low doping concentrations. It has the same selection rule as the inter-miniband absorption (electric field perpendicular to the layers) and an oscillator strength close to unity.

4. Results and discussion

In figure 4 the absorption spectrum for sample no 1 is shown, measured at $T = 5$ K and 300 K. The linehape is essentially independent of temperature except for a somewhat larger line broadening at higher temperature. This is because the Fermi energy at low temperature is

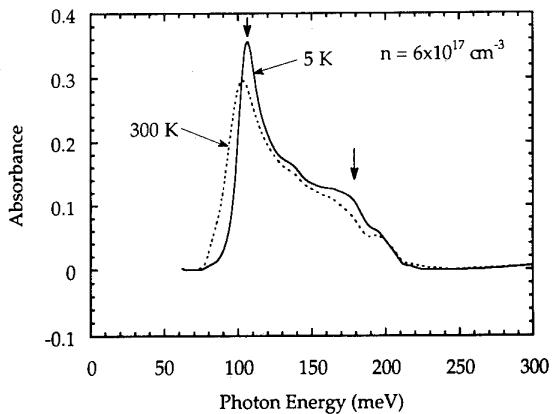


Figure 4. Experimental absorption spectrum of superlattice no 1 ($n = 6 \times 10^{17} \text{ cm}^{-3}$) at $T = 5 \text{ K}$ and $T = 300 \text{ K}$. The peaks resulting from the critical points at $k_z = 0$ and $k_z = \pi/d$ are indicated. (Absorbance = $-\log_{10}$ (transmission).)

well above the first miniband, so that the thermal smearing of the distribution function cannot change the relative occupation of the miniband edges significantly. Both critical points of the superlattice Brillouin zone are clearly revealed, and we find it remarkable that even the relative height of the two absorption maxima is well described by the above calculation.

In the lower-doped sample no 2 the situation is qualitatively different. Since at low temperatures the Fermi energy lies approximately in the middle of the first miniband, the empty zone-edge states can be populated by increasing the temperature. This leads to a strong temperature dependence of the absorption spectrum, as can be observed in figure 5. The large low-energy absorption peak, which is due to the zone-edge transition, can be switched on and off with raising or lowering the temperature respectively. This behaviour is a direct manifestation of the k -space dispersion of the electrons in the first miniband. In addition, a third peak becomes visible at low temperatures, which can be identified with the $1s-2p_z$ donor transition, i.e. with the transition between the localized states at the bottom of each miniband.

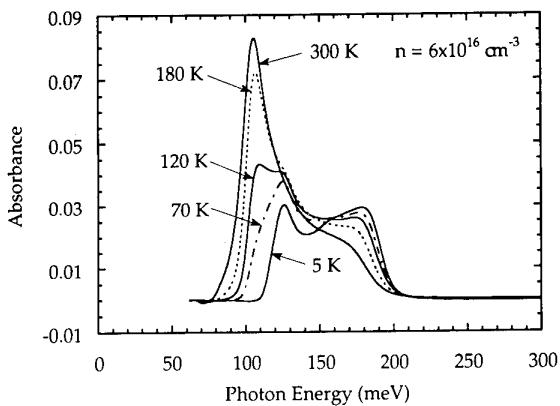


Figure 5. Experimental absorption spectrum of superlattice no 2 ($n = 6 \times 10^{16} \text{ cm}^{-3}$) at different temperatures as indicated.

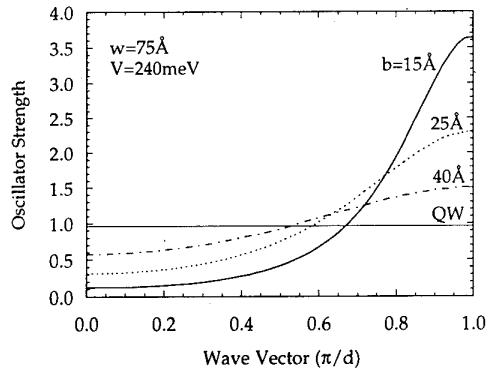


Figure 6. Oscillator strength f_{12} as a function of k_z for GaAs/Al_{0.3}Ga_{0.7}As superlattices with the same well width (75 Å) but different barrier widths, b . The corresponding widths of the first miniband are $\Delta_1 = 36 \text{ meV}$, 18 meV (present samples), and 7 meV, for $b = 15$, 25 and 40 Å respectively. The horizontal line represents the case of an infinite quantum well.

Comparing the theoretical data (figures 1 and 3) with the experimental data there is a slight disagreement over the exact peak positions. This is most likely due to the fact that the accurate well-to-barrier thickness ratio and the Al content in the barriers are not known exactly, as well as the influence of non-parabolicity, which will certainly affect the higher-energy states.

As mentioned above, the strong asymmetry of the inter-miniband absorption line results from the variation of the oscillator strength along the superlattice Brillouin zone. This is illustrated in figure 6, where the oscillator strength is plotted as a function of k_z for superlattices with the same well width and barrier height, but different barrier widths as indicated. Evidently, for thin barriers (or wide minibands) most of the oscillator strength is concentrated near the zone edge. In the opposing limit of very narrow minibands (or dispersionless two-dimensional subbands) the oscillator strength is independent of k_z and of the order unity ($f_{12} = 0.96$ for an infinite potential well, also shown in figure 6). Note, however, that the average of $f_{12}(k_z)$ across the mini-zone remains of the order unity even for wide bands [17].

As clearly seen in figure 5, the k_z dependence of f_{12} leads to an apparent non-conservation of the integrated absorption as a function of temperature. The solution to this puzzling observation lies in the existence of Drude-like *intra*-miniband absorption, which is due to impurity- or phonon-assisted transitions. This one-dimensional Drude absorption has to contain the area that is missing in the low-temperature spectra in figure 5. Even though this absorption could not be directly observed in the present experiment, since it occurs at much lower photon energies (< 20 meV), it has in fact been experimentally observed previously [19]. Obviously, such transitions are only possible in a partially filled band as long as kT is smaller than the bandwidth. Owing to the existence of such *intraband* transitions, the oscillator strength has to be smaller than unity at the zone centre, and larger at the zone edge. This is necessary in order to fulfil the fundamental oscillator sum rule (*f*-sum rule) for band

electrons comprising both *inter-* and *intra*-miniband transitions [17, 20].

Reconsidering the present absorption spectra, it is plausible that the competition between zone-centre and zone-edge absorption will depend on the *relative* position of E_F with respect to the miniband structure. In contrast, the competition between miniband and impurity transitions will be governed by the *absolute* doping level. In our experiment, both samples were metallic at low temperatures, since they are doped well above the Mott criterion for the metal-insulator transition (MIT). The present kind of spectroscopy permits a determination of the position of E_F . For instance, even in the low-doped sample E_F is located in the conduction band at low temperatures, since the zone-centre transition persists down to the lowest temperatures. It will be interesting to investigate the change of the absorption spectrum when a MIT is induced by a magnetic field. According to current knowledge from experiments in bulk GaAs [21], E_F should be located in the impurity band at the MIT. For superlattices, which on the one hand represent a strongly anisotropic system and on the other hand exhibit layer thickness fluctuations as an additional source of localization, the situation is not yet known. With regards to the higher-doped sample, it cannot be excluded that the impurity transition persists at this doping level, but is masked by the strong miniband zone-edge absorption in our experiment. Experiments on samples with wider bands will cast light on this issue.

5. Conclusions

We have demonstrated that infrared spectroscopy can reveal some key features of the band structure of semiconductor superlattices, as there is miniband dispersion together with the van Hove singularities at the zone boundaries. The lineshape can be understood by considering the oscillator strengths and the fundamental optical sum rule. Furthermore, the interplay between extended and localized states can be studied. In future investigations this will be extended to the study of the metal-insulator transition. Finally, the investigation of the infrared absorption spectrum under the conditions of Bloch oscillations and the Wannier-Stark ladder remains a challenge for the future.

Acknowledgments

This work has been supported by the ‘Fonds zur Förderung der wissenschaftlichen Forschung’ (FWF), Vienna, by the Texas Advanced Research Program and

by the Texas Advanced Technology Program. One of us (FMP) acknowledges support by the Belgian National Science Foundation.

References

- [1] Esaki L and Tsu R 1970 *IBM J. Res. Dev.* **14** 61
- [2] Sibille A, Palmier J F, Wang H and Mollot F 1990 *Phys. Rev. Lett.* **64** 52; Sibille A, Palmier J F and Mollot F 1992 *Appl. Phys. Lett.* **60** 457
- [3] Beltram F, Capasso F, Sivco D L, Hutchinson A L, Chu S-N G and Cho A Y 1990 *Phys. Rev. Lett.* **64** 3167
- [4] Grahn H T, von Klitzing K, Ploog K and Döhler G H 1991 *Phys. Rev. B* **43** 12094
- [5] Lee M, Solin S A and Hines D R 1993 *Phys. Rev. B* **48** 11921
- [6] Waschke Ch, Roskos H G, Schwedler R, Leo K, Kurz H and Köhler K 1993 *Phys. Rev. Lett.* **70** 3319
Leo K, Haring Bolivar P, Brüggemann F, Schwedler R and Köhler K 1992 *Solid State Commun.* **84** 943
Feldmann J *et al* 1992 *Phys. Rev. B* **46** 7252
- [7] Mendez E E, Agullo-Rueda F and Hong J M 1988 *Phys. Rev. Lett.* **60** 2426
Voisin P, Bleuse J, Bouche C, Gaillard S, Alibert C and Regreny A 1988 *Phys. Rev. Lett.* **61** 1639
- [8] Lane P and Greene R L 1986 *Phys. Rev. B* **33** 5871
- [9] Helm M, Peeters F M, DeRosa F, Colas E, Harbison J P and Florez L T 1992 *Surf. Sci.* **263** 518
- [10] Song J J *et al* 1989 *Phys. Rev. B* **39** 5562
Deveaud B *et al* 1990 *Phys. Rev. B* **40** 5802
Moore K, Duggan G, Raukema A and Woodbridge K 1990 *Phys. Rev. B* **42** 1326
Fujiwara K, Kawashima K, Yamamoto T, Sano N, Cingolani R, Grahn H T and Ploog K 1994 *Phys. Rev. B* **49** 1809
- [11] Helm M, Hilber W, Fromherz T, Peeters F M, Alavi K and Pathak R N 1993 *Phys. Rev. B* **48** 1601
- [12] Greene R L and Bajaj K K 1985 *Phys. Rev. B* **31** 4006
- [13] Helm M, Peeters F M, DeRosa F, Colas E, Harbison J P and Florez L T 1991 *Phys. Rev. B* **43** 13983
- [14] Hertle H, Schuberth G, Gornik E, Abstreiter G and Schäffler F 1991 *Appl. Phys. Lett.* **59** 2977
- [15] Ihm G, Falk M L, Noh S K, Lee S J and Kim T W 1992 *Superlatt. Microstruct.* **12** 159
- [16] As an example of a more advanced theory see, for example, Serre J and Ghazali A 1982 *Phys. Rev. B* **28** 4704
- [17] Matulis A, Peeters F M, Helm M, Fromherz T and Hilber W 1993 *Phys. Rev. B* **48** 12008
- [18] Karunasiri R P G and Wang K L 1988 *Superlatt. Microstruct.* **4** 661
- [19] Brozak G, Helm M, DeRosa F, Perry C H, Koza M, Bhat R and Allen S J Jr 1990 *Phys. Rev. Lett.* **64** 3163
- [20] Wooten F 1972 *Optical Properties of Solids* (New York: Academic)
- [21] Ehrenreich H and Cohen M H 1959 *Phys. Rev.* **115** 786
Lee M-W, Romero D, Drew H D, Shayegan M and Elman B S 1988 *Solid State Commun.* **66** 23
Romero D, Liu S, Drew H D and Ploog K 1990 *Phys. Rev. B* **42** 3179

Wannier–Stark effect in superlattices

C Hamaguchi†, M Yamaguchi†, M Morifujit, H Kubo†,
K Taniguchi†, C Gmachl‡ and E Gornik‡

† Department of Electronic Engineering, Faculty of Engineering, Osaka University,

Suita City, Osaka 565, Japan

‡ Institut für Festkörperelektronik, Technische Universität Wien, Gusshausstraße

25–29/362, A-1040 Wien, Austria

Abstract. Electroreflectance measurements have been carried out in order to investigate Stark-ladder transitions in a GaAs (40 Å)/AlGaAs (20 Å) superlattice under various uniform electric fields, and compared with the transition energies calculated on the basis of a microscopic tight-binding theory. The observed electroreflectance spectra over a wide range of photon energies (1.5–2.2 eV) shift in proportion to an applied electric field. The signals in a higher photon energy region (1.9–2.2 eV) indicate the existence of a transition from the spin-orbit split-off band in the valence band to the Wannier–Stark localization states in the conduction band. The assignment is supported by the tight-binding calculation. Resonant coupling between the localized states is also observed.

1. Introduction

The electronic properties of heterostructures have received great interest because of their potential application to high-speed electron devices and optoelectronic devices. In the heterostructures, a superlattice (SL), consisting of a periodic repetition of quantum wells separated by narrow potential barriers, has been investigated mainly with regard to its optical properties.

In the SL, the electronic wavefunctions are completely delocalized and their energies are distributed in minibands due to the strong coupling between the wells. Under this condition the interband optical transition between conduction and valence bands will be allowed at each position of the SL. The behaviour of electrons and holes in a uniform electric field in such a SL has recently attracted attention [1–4]. In a high electric field, the wavefunctions will be localized in each quantum well, which is called Wannier–Stark (WS) localization, and the interband transitions will be restricted to these regions. When a low uniform electric field is applied to the SL, the energy degeneracy between the wells is broken, resulting in discrete levels. Therefore it is possible to observe new transitions, which are called Stark-ladder transitions.

A series of these phenomena is called the WS effect, and the dependence of the Stark-ladder transition energies on the applied electric field is the characteristic of WS effect. The Stark-ladder transition energies shift in proportion to an applied electric field F , which is approximated by

$$E_v = E + veFd \quad (v = 0, \pm 1, \pm 2, \dots) \quad (1)$$

where E_v are transition energies, E is an interband transition energy, v is the Stark-ladder index, e is the

electronic charge and d is the SL period. The magnitude of the Stark shifts depends on the second term of equation (1).

The WS localization has been investigated by observing photocurrent, optical transitions and so on [2, 4, 5]. Modulation spectroscopy, such as the electroreflectance (ER) and photoreflectance (PR) methods, is known to be the most powerful and precise method for obtaining the optical transitions in SLs. In our previous works, we carried out PR measurements in GaAs/AlAs short-period SLs in order to clarify the energy band structures and found the existence of a quasi-direct transition due to the zone folding effect of the Brillouin zone [6, 7]. No external electric field was applied in the experiment. In the present work we are interested in the WS effect and thus we performed the ER experiments in the presence of an external electric field, where we investigated the optical transitions in a higher photon energy region in addition to the region near the fundamental absorption edge in the GaAs/AlGaAs superlattices. These experimental results were compared with the transition energies calculated by the tight-binding (TB) theory, in which atomic-like wavefunctions are used as the basis set. The TB calculations include the effect of the electric field in the diagonal elements of the Hamiltonian matrix, and take account of the intraatomic spin-orbit interaction [8]. From the comparison between the experimental results and the calculations, we found that the observed transition energies in the low photon energy region show good agreement with the calculations and that the observed ER signals at higher photon energy arise from the localized state in the spin-orbit split-off band.

2. Experiments

The experiments were carried out for a GaAs (40 Å)/Al_{0.3}Ga_{0.7}As (20 Å) SL structure grown on an n⁺-GaAs substrate by molecular beam epitaxy. The SL has 50 periods and is cladded by undoped GaAs (60 Å) and undoped Al_{0.3}Ga_{0.7}As (60 Å) spacers. These layers are placed in a p-i-n diode structure, where the n- and p-layers are 0.6 μm thick n-GaAs (1 × 10¹⁸ cm⁻³) and 0.4 μm thick p-Al_{0.3}Ga_{0.7}As (5 × 10¹⁷ cm⁻³). Cap layers of 0.2 μm thick p⁺-Al_{0.3}Ga_{0.7}As (1 × 10¹⁹ cm⁻³) are also grown.

We have studied ER spectroscopy of this GaAs/AlGaAs SL at given uniform electric fields. The uniform electric field in the ER measurement is modulated by a small DC field produced by applying a small AC voltage of amplitude 100 mV and frequency 1 kHz. The probe light is monochromatic light from a tungsten lamp dispersed with a 50 cm single monochromator. The probe light reflected from the sample is detected by a photomultiplier tube and amplified by a lock-in amplifier. The sample temperature was 77 K.

Figures 1(a) and (b) show the observed ER spectra of the GaAs/AlGaAs SL at various applied voltages (from +0.6 V to -3.0 V). Spectra rich in structures are observed over a wide range of photon energies (1.5–2.2 eV). We did not try to fit the lineshape of the third-derivative formula to the experimental data [9, 10], because the observed reflectance signals of the Stark-ladder transitions are quite complicated, as shown in figures 1(a) and (b). However, the extrema in the form of derivative spectra give rise to an accurate determination of the critical point energies. Thus we regarded extrema in the ER signals as critical point energies. The reflectance signals shown in figure 1(a) shift in proportion to the applied voltage towards both

lower and higher energies, and we find that the transition energies are well expressed by equation (1). Figure 1(a) illustrates the observed signals of H₁E₁ ν and L₁E₁ ν ($\nu = 0, \pm 1$ and ± 2) in the photon energy region 1.5–1.8 eV. The labels H_mE_n ν (L_mE_n ν) indicate transitions with the Stark-ladder index ν between the m th heavy-hole (light-hole) subband and the n th electron subband. The ER signals in the photon energy region 1.8–2.2 eV are shown in figure 1(b), which will be discussed later.

3. Energy band calculation

The ER spectra over a wide range of energies clearly indicate the existence of Stark-ladder transitions between the valence band and conduction band states. In order to analyse these spectra, especially in the higher photon energy region, we extend the empirical TB theory used for the energy band calculation [6, 7, 11, 12], to the case of the SL in an electric field.

In our previous paper we reported that the experimental data from the photoluminescence and PR measurements in GaAs/AlAs SLs without electric field are well explained by the energy band structure calculated by the TB method [6, 7]. In the calculations basis functions are taken as a three-dimensional Fourier transformation of atomic orbitals. These basis functions are labelled by three-dimensional wavevectors. On the other hand, when an electric field is applied along the growth axis (z-axis), the translational symmetry in the direction of the electric field is lost. In such a case we are not able to apply the Fourier analysis in the z-direction, although it is still valid in the x- and y-directions. Therefore, the basis functions are specified by the atomic position l along the z-axis, and a two-dimensional wavevector \mathbf{k} as

$$\psi_{\mathbf{k},l}^{b\mu}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{i,j} \exp(i\mathbf{k} \cdot \mathbf{r}_{i,j}) \phi_{i,j,l}^{b\mu}(\mathbf{r}) \quad (2)$$

where $\phi_{i,j,l}^{b\mu}(\mathbf{r})$ is an atomic orbital located at the site of the crystal (i, j, l) and b and μ denote the kind of atom ($b = \text{cation (c)} \text{ or anion (a)}$) and the type of atomic orbital ($\mu = s, p_x, p_y \text{ or } p_z$). Considering that the optical properties are mainly governed by the band structure at the Γ -point, we set $\mathbf{k} = 0$.

We calculate energies and wavefunctions of WS localization states by solving the secular equation

$$\det|H_{lb\mu;l'b'\mu'} - \delta_{l,l'}\delta_{b,b'}\delta_{\mu,\mu'}\epsilon| = 0 \quad (3)$$

where

$$H_{lb\mu;l'b'\mu'} = \langle \psi_l^{b\mu}(\mathbf{r}) | H | \psi_{l'}^{b'\mu'}(\mathbf{r}) \rangle. \quad (4)$$

The effect of the electric field is included in the diagonal elements of the Hamiltonian matrix, and the Hamiltonian contains the intra-atomic spin-orbit interaction. We have neglected a contribution from field-induced intra-atomic mixing between s and p_z orbitals, since it is expected to be very small.

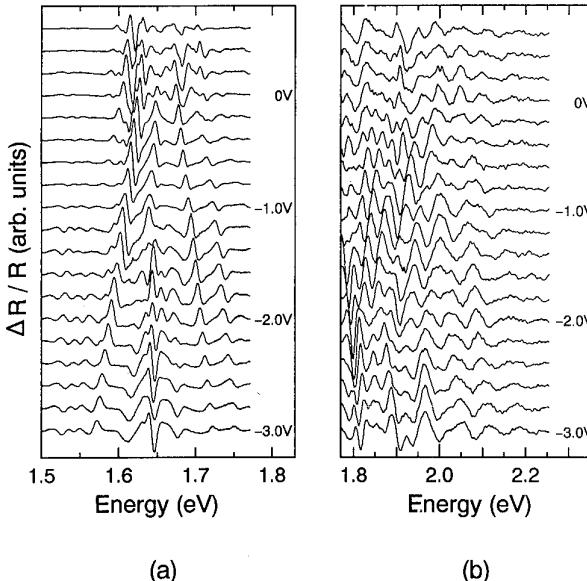


Figure 1. Electroreflectance spectra for the GaAs/Al_{0.3}Ga_{0.7}As superlattice observed at various applied voltages and at 77 K, for DC biases from +0.6 V to -3.0 V in steps of 0.2 V as indicated.

In order to solve the secular equation, equation (3), we consider a finite domain in the z -direction, because the number of SL periods of a real system is finite and no translational symmetry exists along the z -direction. For AlGaAs layers we used a virtual crystal approximation in which the TB parameters are replaced by an average value of GaAs and AlAs. In the present study we calculate the energy band structure for ten periods of $(\text{GaAs})_{14}/(\text{Al}_{0.3}\text{Ga}_{0.7}\text{As})_7$. The valence band offset was set to be 0.144 eV.

4. Discussion

From the TB calculation we find that the heavy-hole states are well localized in each GaAs quantum well at 30 kV cm $^{-1}$ and that the conduction band states are not completely localized at this electric field, reflecting their lighter effective mass. However, the TB calculation shows that the conduction band states are localized in a higher electric field [13].

The ER spectra in the higher photon energy region (1.9–2.2 eV; figure 1(b)) may suggest the localization of spin-orbit split-off (SO) states due to an electric field. However, the existence of the Stark ladder of SO states is not apparent, since the SO states have energies larger than the barrier in the valence band of the SL. In addition, higher minibands of heavy and light holes which are folded into the first Brillouin zone are close to the SO states in terms of energy, and mixing between the states of these minibands and the SO states may delocalize the Stark ladder of the SO states.

We calculated the charge densities of an SO state at several electric field strengths in order to clarify whether SO states are localized. Calculated charge densities of an SO eigenstate are presented in figure 2, where the squared amplitude of the state is plotted at each atomic site for several electric fields. This figure clearly shows that the SO states are gradually confined into a GaAs layer as the electric field increases, indicating the existence

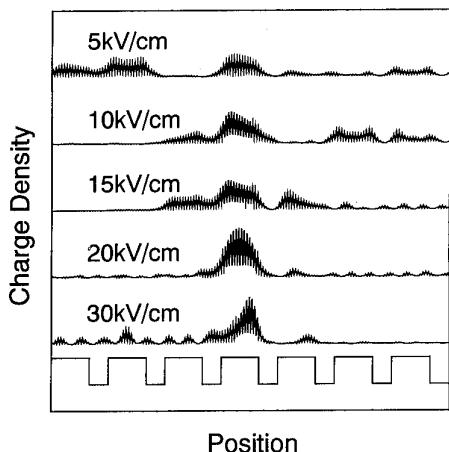


Figure 2. Charge density distribution of the spin-orbit split-off state calculated by the tight-binding method.

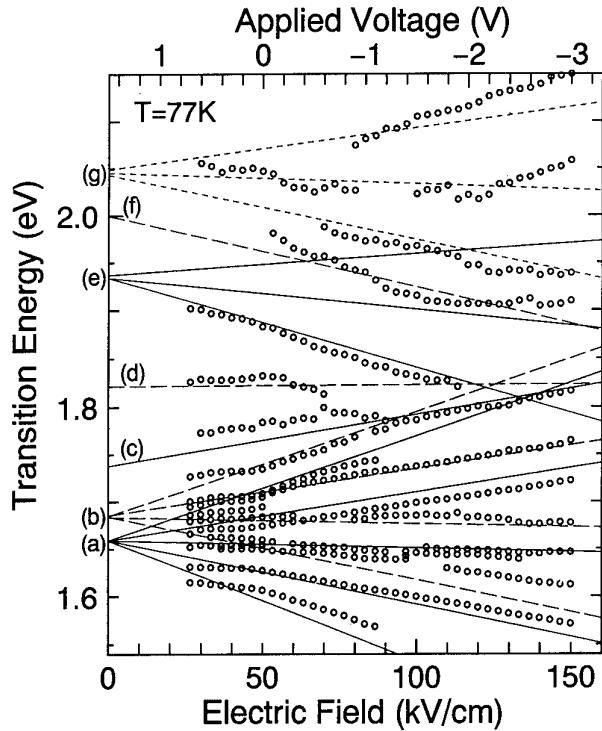


Figure 3. Calculated transition energies from valence band states to conduction band states. The full curves (a) in the lower energy region correspond to the transitions H_1E_{10} , $H_1E_1 \pm 1$ and $H_1E_1 \pm 2$. The broken curves (b) in the lower energy region correspond to the transitions L_1E_{10} , $L_1E_1 \pm 1$ and $L_1E_1 + 2$. The full curve (c) extrapolated to 1.74 eV is $H_2E_1 + 1$, and the broken curve (d) at 1.82 eV corresponds to L_2E_{10} . The full curves (e) in the higher energy region correspond to the transitions H_2E_{20} and $H_2E_2 \pm 1$, and the broken curve (f) extrapolated to 2.0 eV is $L_2E_2 - 1$. The dotted curves (g) show the transition energies from the localized spin-orbit split-off states to the conduction band states. The observed extrema of the electroreflectance spectra are shown by open circles. See the text for the notation $H_mE_n\nu$ and $L_mE_n\nu$.

of a Stark ladder of SO states. The charge density at $F = 30$ kV cm $^{-1}$ is delocalized and has a long tail that extends over many SL periods. This delocalization is due to the mixing between a localized SO state and an extended state of a miniband which is folded into the first Brillouin zone. Since the densities of the extended states in terms of energy become larger as the size of the system (number of SL periods) increases, the Stark ladder of the SO state will vanish after an infinite number of periods of an SL. Real systems, however, are finite and thus the Stark ladder is expected to exist.

Figure 3 shows the calculated transition energies between valence band states and conduction band states as a function of the electric field, where the observed extrema of the ER spectra are also shown by open circles. Experimental values of the electric fields are calculated from the built-in voltage of 1.50 V, which is obtained by extrapolating the relationship between ER extrema and applied voltages. We note that resonant coupling between neighbouring localized states that occur at crossing points of transition energies is neglected in this figure, and this will be discussed later. The full

curves (a) in the lower energy region correspond to the transitions H_1E_10 , $H_1E_1 \pm 1$ and $H_1E_1 \pm 2$, while the full curve (c) extrapolated to 1.74 eV is $H_2E_1 + 1$, and the full curves (e) in the higher energy region correspond to the transitions H_2E_20 and $H_2E_2 \pm 1$. The broken curves (b) in the lower energy region correspond to the transitions L_1E_10 , $L_1E_1 \pm 1$ and $L_1E_1 + 2$. The broken curve (d) at 1.82 eV is the transition corresponding to L_2E_10 , and the broken curve (f) extrapolated to 2.0 eV is $L_2E_2 - 1$.

The good agreement between the observation and the calculation indicates that the Stark ladder is well described within the framework of the TB theory. In particular, the signals at lower photon energies (1.5–1.8 eV) are well explained as the Stark-ladder transitions between the valence band states and the conduction band states. The calculated transition energies between localized SO states and the localized conduction band states are shown in this figure by dotted curves (g). Although there is a slight deviation, the calculated curves fit the observed ER signals in the photon energy region 1.9–2.2 eV. We also found that the calculated optical transition probabilities between SO states and conduction band states have a magnitude large enough to be observed. This agreement indicates that the observed ER spectra in the higher photon energy region are attributed to the Stark-ladder transitions between the SO states and the conduction band states.

Next, we will discuss resonant coupling of the Stark-ladder states [14–16]. Figures 4(a) and (b) give full details of the ER spectra around the photon energy 1.64 eV at applied voltages from -1.06 V to -1.70 V and from -2.60 V to -2.98 V in steps of 20 mV. Both figures show that the two kinds of transition energy do not cross each other, apparently indicating the existence of resonant coupling. Resonant coupling between neighbouring localized states is expected to occur at where their transition energies cross. In figure 3 we plotted the calculated transition energies by neglecting the splitting induced by resonant coupling. However, the ER spectra in figures 4(a) and (b) indicate the existence of resonant coupling between neighbouring localized states. The TB calculation also indicates that the ER spectra in figures 4(a) and (b) are due to the resonant coupling between H_1E_10 and $H_2E_1 - 2$ and between H_1E_10 and $H_2E_1 - 1$, respectively. The energies of the minima in the ER spectra in figure 4(b) are plotted as full circles as a function of the electric field in figure 5, where the calculated transition energies of both H_1E_10 and $H_2E_1 - 1$ are also plotted by full curves, and the broken curves are the transition energies without coupling between the states obtained by extrapolating H_1E_10 and $H_2E_1 - 1$ at lower electric fields. We found that the H_1E_10 transition energy and the $H_2E_1 - 1$ transition energy do not cross each other either experimentally or theoretically, and that the experimental data agree well with the calculated results. These results indicate that the resonant coupling between H_1E_10 and $H_2E_1 - 1$ occurs at 142 kV cm⁻¹ (at -2.76 V).

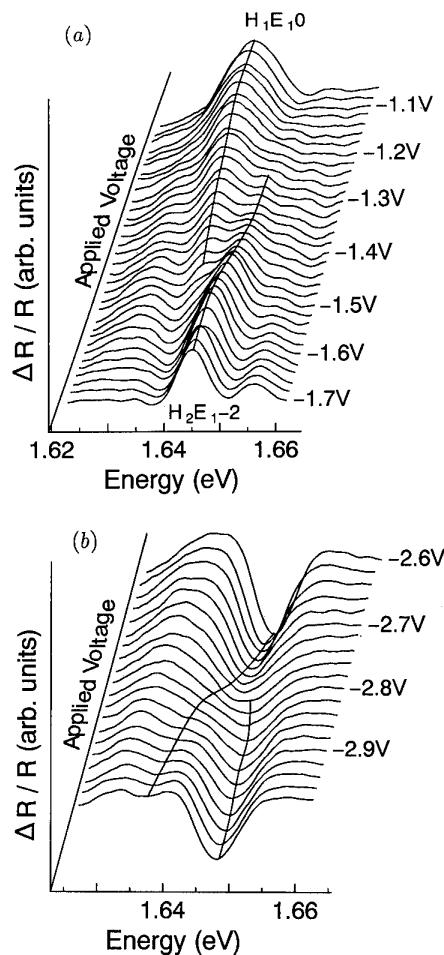


Figure 4. Electroreflectance spectra around the photon energy 1.64 eV for the GaAs/Al_{0.5}Ga_{0.7}As superlattice at applied voltages (a) from -1.06 V to -1.70 V and (b) from -2.60 V to -2.98 V in steps of 20 mV as indicated, at 77 K.

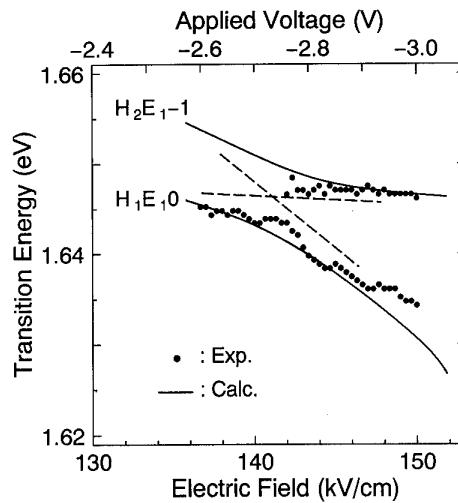


Figure 5. Comparison of the energies of the minima in the electroreflectance spectra in figure 4(b) with the calculated results. Full circles are the observed electroreflectance signals, and the full curves are the calculated transition energies for H_1E_10 and $H_2E_1 - 1$. The broken curves are extrapolated values neglecting the resonant coupling.

5. Conclusion

We summarize the present results as follows. We made ER measurements of a GaAs (40 Å)/AlGaAs (20 Å) SL. The observed extrema in the ER spectra shift as the electric field applied along the growth axis of the SL increases, and they clearly indicate the existence of the Stark ladder, i.e. electric field-induced localization states. In addition, we also observed ER signals in a higher photon energy region, which suggests the existence of a Stark ladder of SO states. In order to analyse the ER spectra, we carried out an energy band calculation based on the TB theory including the intra-atomic spin-orbit interaction. By comparing the calculated results with the observed data, we found that the ER spectra in the lower photon energy region agree well with the calculated results and thus that the Stark ladder is well described by the TB theory. Furthermore, we found that the SO states are localized in an electric field as well as the states at the band edge. The calculated transition energies from the localized SO states to the localized conduction band states agree well with the ER spectra in the energy region 1.9–2.2 eV. We also observed resonant coupling between H_1E_10 and $H_2E_1 - 2$ and between H_1E_10 and $H_2E_1 - 1$. The anti-crossing of two resonant couplings is seen clearly in the ER measurements (figures 4(a) and (b)), and the experimental data agree well with the calculated results.

References

- [1] Bleuse J, Bastard G and Voisin P 1988 *Phys. Rev. Lett.* **60** 220
- [2] Mendez E E, Agulló-Rueda F and Hong J M 1988 *Phys. Rev. Lett.* **60** 2426
- [3] Voisin P, Bleuse J, Bouche C, Gaillard S, Alibert C and Regreny A 1988 *Phys. Rev. Lett.* **61** 1639
- [4] Agulló-Rueda F, Mendez E E and Hong J M 1989 *Phys. Rev. B* **40** 1357
- [5] Fujiwara K 1989 *Jpn. J. Appl. Phys.* **28** L1718
- [6] Fujimoto H, Hamaguchi C, Nakazawa T, Taniguchi K, Imanishi K, Kato H and Watanabe Y 1990 *Phys. Rev. B* **41** 7593
- [7] Matsuoka T, Nakazawa T, Ohya T, Taniguchi K, Hamaguchi C, Kato H and Watanabe Y 1991 *Phys. Rev. B* **43** 11798
- [8] Morifumi M, Nishikawa Y, Hamaguchi C and Fujii T 1992 *Semicond. Sci. Technol.* **7** 1047
- [9] Aspnes D E and Rowe J E 1972 *Phys. Rev. B* **5** 4022
- [10] Aspnes D E 1973 *Surf. Sci.* **37** 418
- [11] Chadi D J 1977 *Phys. Rev. B* **16** 790
- [12] Vogl P, Hajimerson P and Dow J D 1983 *J. Phys. Chem. Solids* **44** 365
- [13] Yamaguchi M, Morifumi M, Kubo H, Taniguchi K, Hamaguchi C, Gmachl C and Gornik E 1993 *Proc. SPIE* **1985** 608
- [14] Schneider H, Grahn H T, von Klitzing K and Ploog K 1990 *Phys. Rev. Lett.* **65** 2720
- [15] Fox A M, Miller D A B, Livescu G, Cunningham J E and Jan W Y 1991 *Phys. Rev. B* **44** 6231
- [16] Tanaka I, Nakayama M, Nishimura H, Kawashima K and Fujiwara K 1992 *Phys. Rev. B* **46** 7656

Optical spectroscopy on quantum wells and tunnel structures under conservation of angular momentum

G Hendorfer

Institut für Experimentalphysik, Johannes-Kepler-Universität, Altenberger Strasse 69, A-4040 Linz, Austria

Abstract. We show that the use of circularly polarized light for the excitation in optical experiments (optical pumping) such as excitation spectroscopy or photoluminescence gives a better optical resolution as well as a much more reliable interpretation of the results when confined systems are investigated. We present an investigation of pseudomorphic strained InGaAs/GaAs and InGaAs/AlGaAs structures by means of optical pumping and give the conduction band offsets obtained from these experiments. We show that by applying a transverse magnetic field the circular polarization of the luminescence is decreased. By analysing such experiments in combination with time-resolved spectroscopy we determine the excitonic spin relaxation times and the excitonic *g*-factors. From these results we see that details of the confinement and strain conditions enter into these quantities. Finally we demonstrate that the spin of electrons is conserved during the tunnelling process through the barriers of tunnel structures when the barrier thickness is below a critical value. From our results we conclude that the interface quality determines the spin relaxation in heterostructures much more strongly than any other property.

1. Introduction

The use of circularly polarized light for the excitation in optical experiments like photoluminescence (PL) or photoluminescence excitation spectroscopy (PLE) is known as optical pumping (OP) in semiconductor physics [1]. This technique was developed in atomic physics [2] and has been used by French and Russian groups to investigate bulk crystals [3]. Later the method was applied to investigate low-dimensional systems [4–7]. In optical pumping experiments the angular momentum of light is transferred to free states of a semiconductor. Thus spin-polarized carriers are created. The hole spins are supposed to relax much faster than the electron spins because their spin relaxation time is coupled via the spin-orbit interaction to the hole momentum relaxation time. So only spin-polarized electrons can survive the relaxation of the free carriers to their respective ground states. In low-dimensional systems, the hole spin relaxation times can be much longer than in bulk samples [8, 9]. They are, however, much shorter (1 ps or less) than typical electron spin relaxation times. Thus the above assumption that mainly electrons are polarized holds also in undoped quantum wells. These electron spins are oriented along the *k* vector of the incident light. If the rate of the optical decay is higher than the rate of electronic spin-flip processes these electronic spins are transferred to the light of the photoluminescence thus creating circularly polarized light.

We have investigated pseudomorphic strained InGaAs/GaAs and InGaAs/AlGaAs quantum wells grown on a (001) substrate. In such structures the built-in strain offers an additional degree of freedom for tailoring specific devices. On the other hand the strain also modifies important band structure parameters like effective masses [10] and band offsets, which changes the whole subband structure. This makes its assessment via excitonic peaks more difficult than in the case of lattice-matched structures. In particular, discrimination between heavy-hole (hh) and light-hole (lh) excitons can be difficult, and reliable assignments can rarely be obtained by comparison with model calculation alone. Progress towards identification of lh excitons has been made for strained InGaAs/GaAs wells in absorption [11] and PLE [12] measurements using light polarized linearly within and perpendicular to the QW layer. However, since QWs (quantum wells) have axial symmetry around the growth direction circularly polarized light is ideally suited to probe such structures.

We show that by OP it is possible to provide a unique means of discriminating between hh and lh excitons in the above-mentioned structures. We have combined OP with depolarization measurement in a transverse magnetic field to obtain information on spin relaxation times and *g*-factors. These quantities, in turn, very critically depend on the confinement and strain conditions. From our experiments we determine also the conduction band offsets (under strain conditions) of both InGaAs/GaAs and InGaAs/AlGaAs heterostructures.

Additionally we have investigated the spin conservation of electrons during their non-resonant tunnelling through the barriers of asymmetric GaAs/AlGaAs tunnel structures by optically detected tunnelling. We demonstrate that the spin conservation depends much more strongly on the thickness of the barrier than the normal tunnel effect. This gives the possibility of studying the quality of interfaces.

2. Optical pumping in pseudomorphic layers

The QW structures investigated have been grown on a (001)-oriented GaAs substrate and their conduction band structures are shown schematically in the inset of figures 1(a) and (b). The structures consist of five InGaAs layers having widths of 2, 5, 10, 13 and 16 nm and an In mole fraction of 0.16. They are separated from each other by 30 nm GaAs and Al_{0.3}Ga_{0.7}As layers, and from the substrate by a GaAs/Al_{0.3}Ga_{0.7}As superlattice buffer. The structure is capped by a 30 nm GaAs layer. The widths of the layers have also been checked by transmission electron microscopy and secondary-ion mass spectroscopy. The measured well widths are very close to the nominal ones.

For the optical excitation we have used a 5 W Ar laser and a Ti:sapphire laser which emits tuneable laser light within the wavelength range 0.7 to 1 μm. For the generation of circularly polarized light we have used a photoelastic-modulator which switches the circular

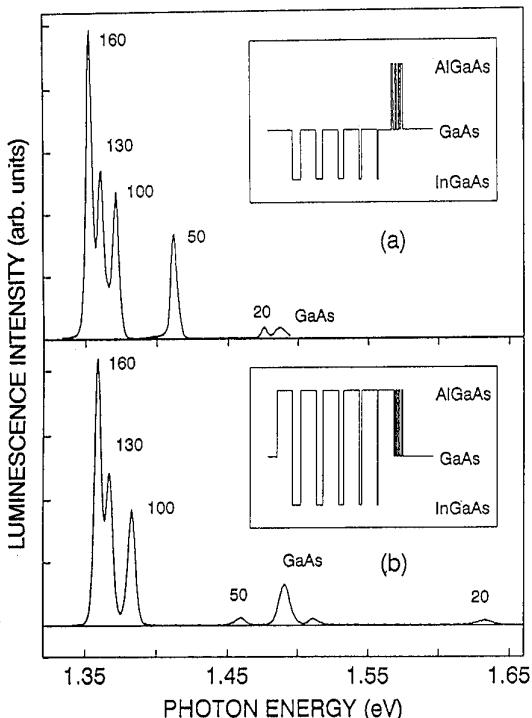


Figure 1. Luminescence intensity versus photon energy for InGaAs/GaAs (a) and InGaAs/AlGaAs (b) quantum wells. The numbers associated with the peaks correspond to the well widths. Excitation power density is 1 W cm⁻².

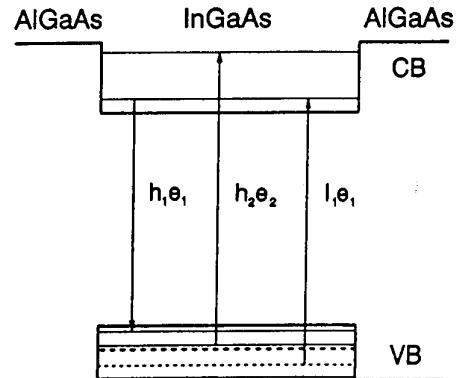


Figure 2. InGaAs/AlGaAs quantum well shown schematically. Band edges are shown as thick lines, quantized levels as thin lines. Band edges and quantized levels referring to the lh band are broken lines.

polarization of the transmitted beam between σ^+ and σ^- polarization. This modulator allows us to use lock-in detection in our measurement. In order to analyse the degree of circular polarization of the luminescence light we have used a quarter-wave plate and a linear polarizer. Detection of the luminescence was done by means of a monochromator and a Ge detector.

The photoluminescence of our structures is shown in figure 1. Five excitonic peaks due to h₁e₁ transitions in the respective wells can be resolved. The widths of the wells are assigned to the respective lines in the figure. The two lines at 1.5 eV are attributed to transitions in the GaAs capping layer.

The principle of a PLE experiment is shown schematically in figure 2. The bold and full lines represent the conduction band and the heavy-hole valence band edge respectively. The broken lines correspond to the light-hole valence band edge. One light-hole state and two heavy-hole states as well as two electron states are drawn in this figure. In PLE experiments the intensity of the h₁e₁ line is recorded as a function of the photon energy of the excited beam. When a resonant transition due to excited states is induced, as is indicated by the l₁e₁ and h₂e₂ transitions in figure 2, the intensity of the e₁h₁ line is increased. The physical quantity that is measured by that experiment is essentially the density of states.

The principle of optical pumping is shown in figure 3. We have assumed the excitation beam to be σ^+ polarized. Then the selection rules due to the conservation of angular momentum allow only two types of transitions which are indicated as full arrows in figure 3. The matrix element of the hh-related transition is three times larger than that of the lh-related transition. This is the reason why in three-dimensional samples, where the hh and lh states are degenerate, the circular polarization of the luminescence cannot exceed 50%. In strained or confined systems, however, the lh states are split off and as a consequence the circular polarization can equal 100% provided the spin relaxation time is long enough. This high electronic polarization can be used to produce highly polarized electrons in photoemission experiments. On the other hand, the split off of the lh states can be exploited

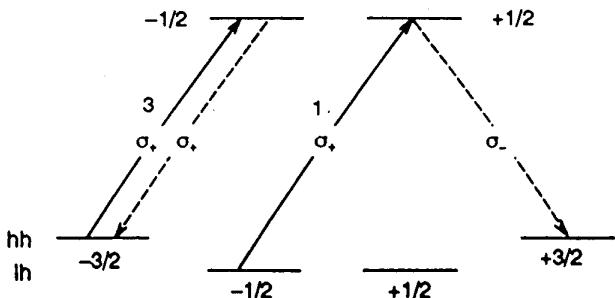


Figure 3. Allowed absorption transitions from the hh and lh valence band to the conduction band (full arrows) under σ^+ polarization. Note that the electron–hh recombination (broken arrows) is σ^+ and σ^- polarized if excitation is from the hh and lh bands respectively.

to discriminate between hh-related and lh-related transitions because the hh-related transitions feed the opposite electronic spin state to lh-related transitions. As figure 3 shows by broken arrows, the two electron spin states recombine due to the selection rules via different channels with heavy holes. Thus resonant excitation of hh excitons are seen as peaks in a PLE experiment whereas resonant excitation of lh excitons is recognized as dips.

An example of this feature is given in figure 4 where the PLE spectra of the 50 Å well of the InGaAs/AlGaAs sample are shown. The upper curve results from the conventional method where the intensity of the emitted light (the sum of the intensity of left- and right-circularly polarized emitted light) is measured. The lower curve,

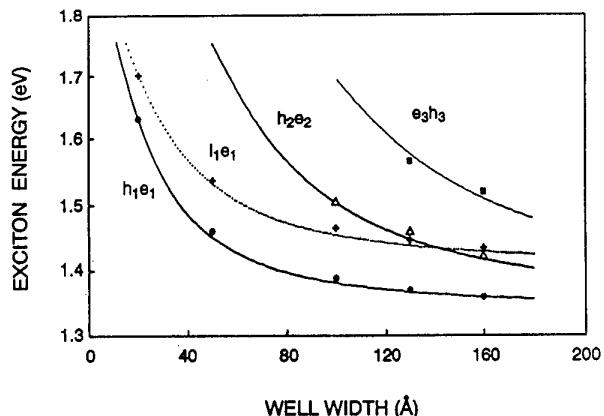


Figure 5. Observed, allowed excitonic transitions (symbols) in the five quantum wells of the InGaAs/AlGaAs structure. The curves are a fit based on a square-well model calculation.

however, results from the optical pumping technique. Here the polarization of the luminescence is plotted versus the photon energy of the polarized exciting light. The $l1e1$ transition shows a clear resonant decrease of the polarization signal. The hh-related transitions manifest themselves as peaks.

We have plotted the exciton binding energy of the allowed transitions (transitions with the same subband index for electrons and holes) versus the InGaAs well width. The plot for InGaAs/AlGaAs is shown in figure 5. The points correspond to experimental results, the full curves correspond to model calculations. The crossover of $l1e1$ and $h1e1$ excitons indicates the presence of strain. This cannot be observed in lattice-matched structures like GaAs/AlGaAs quantum wells. Extrapolating the $l1e1$ and $h1e1$ transitions to infinite well width yields the strain-induced splitting of the lh and hh valence bands in bulk material. From this splitting, and by knowing the deformation potentials of InGaAs, the In mole fraction can be determined independently of line positions. For our model calculation of the exciton energies a square well model is used [13]. An exciton binding energy of 5 meV is supposed in the model calculation. The only fit parameter of this model calculation is the conduction band offset Q . As a result we obtain $Q = 0.70 \pm 0.02$ for $In_{0.16}Ga_{0.84}As/Al_{0.3}Ga_{0.7}As$ and $Q = 0.62$ for $In_{0.16}Ga_{0.84}As/GaAs$. The latter system has been extensively investigated by different methods and the resulting values for Q differ considerably, with values of 0.4 to 0.8 given by different authors [14]. For the former system we give the first value of Q . It is in agreement with a theoretical prediction given by Arent [15].

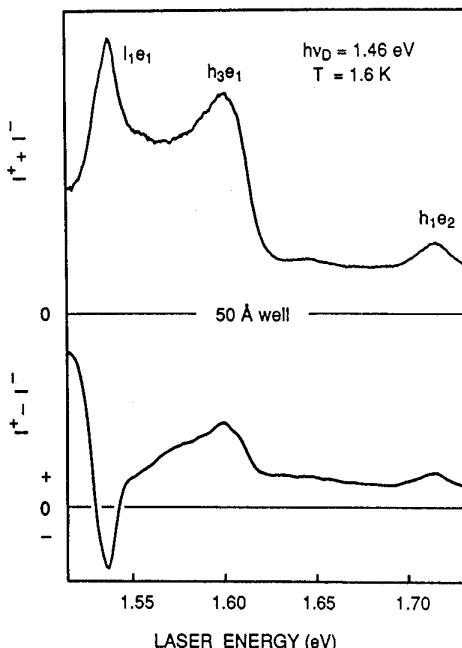


Figure 4. Excitation spectrum of the 50 Å well of InGaAs/AlGaAs above bandgap of the GaAs substrate using σ^+ light. The upper trace shows the total luminescence intensity and the lower trace its degree of circular polarization.

3. Depolarization experiments

The electronic spins oriented due to optical pumping as well as the photoluminescence originating from the

recombination of these electrons can be depolarized by applying a magnetic field normal to the k vector of the incident light. This kind of experiment is also called Hanle depolarization in analogy to the Hanle effect in atomic physics [3]. Before discussing the depolarization effect let us consider the degree ρ of the polarization quantitatively. It is given by $\rho = (1 + \tau/\tau_s)^{-1}$ where τ is the optical decay time of the exciton and τ_s is the spin relaxation time. Thus, measuring ρ allows one to determine the ratio τ/τ_s .

The optical decay time of the excitons in these structures has been determined by time-resolved photoluminescence using picosecond excitation[†]. The obtained decay times are nearly independent of the well width as well as of the barrier material and equal approximately 300 ps. In order to also determine τ_s we have performed magnetic field depolarization experiments.

If a transverse magnetic field is applied to a sample which is optically pumped, the spins precess around the magnetic field axis with a frequency of $\omega = \mu_B gB$. Here μ_B is the Bohr magneton, g the effective g -factor and B the transverse magnetic field. If the effective lifetime T of the excitonic spin state, given by $1/T = (1/\tau) + (1/\tau_s)$ is longer than the time for one period of the spin precession the average spin polarization appears smeared out. Thus a transverse magnetic field reduces the degree of polarization of the luminescence light. The dependence of ρ on B can be calculated by perturbation theory and reads $\rho(B) = (1 + \tau/\tau_s)^{-1}[1 + (\mu_B g B T)^2]^{-1}$. By measuring the polarization at zero field and the half-width of the depolarization curves, and by knowing the optical decay time for each of the h1e1 transitions, we are able to determine the spin relaxation times or spin lifetime as well as the g -factors of the excitons. The results are given in figures 6 and 7. There is a clear tendency towards smaller spin relaxation times with smaller quantum well widths as well as an influence of the barrier material. The sample with the AlGaAs barriers exhibits significantly lower spin lifetimes than the sample with GaAs barriers. We want to point out the sensitivity by which the spin lifetimes react to the specific confinement conditions and suggest the following arguments to explain our results:

(i) The order of magnitude of τ_s (10–50 ps) indicates strong excitonic exchange coupling of the photogenerated carriers.

(ii) The decrease of τ_s with decreasing quantum well width can be understood in terms of the kinetic energy of the carriers.

Stronger confinement implies a higher kinetic energy and as a consequence a higher rate of reflection of the carriers at the interface. Each reflection at the interface exhibits a certain probability for inducing a spin-flip process due to interactions of the carrier spin with paramagnetic defects in the interface region. Therefore the longer spin lifetimes in the sample with the GaAs barriers indicate a better interface quality for InGaAs/GaAs than for

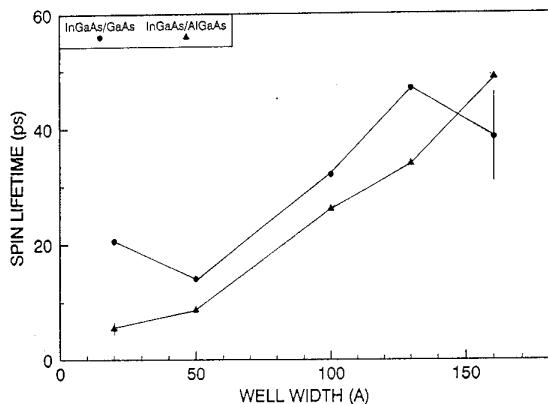


Figure 6. Spin lifetime versus quantum well width of InGaAs/GaAs and InGaAs/AlGaAs.

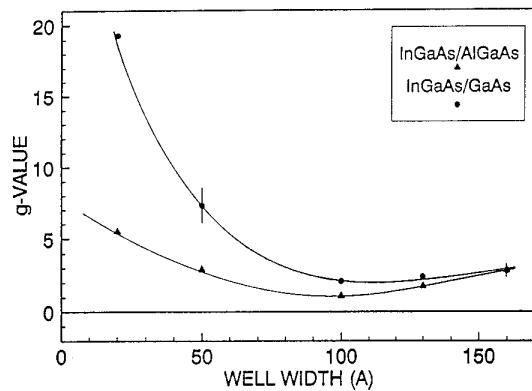


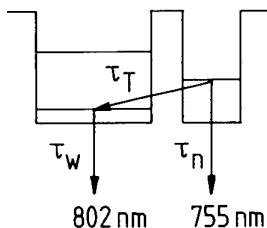
Figure 7. Excitonic g -values versus quantum well width of InGaAs/GaAs and InGaAs/AlGaAs.

InGaAs/AlGaAs. In the latter case a contribution of barrier alloy scattering to the spin-flip rate may be an additional reason for the strongly reduced spin lifetime in the thinner wells. The excitonic g -values exhibit a strong confinement effect, too, as can be seen in figure 7. In this case, however, the confinement effect is stronger for the InGaAs/GaAs quantum wells.

4. Optical pumping in tunnel experiments

The structure investigated in these experiments is an asymmetric double quantum well of GaAs separated by AlGaAs with an Al mole fraction of 0.3. The structure is given in figure 8. The width of the larger well is 10 nm and that of the smaller well is 5 nm. We have used samples with barrier widths of 3, 4, 6, 8 and 20 nm. These samples have been investigated by Nido *et al* [16] with respect to the dynamic properties of the photogenerated carriers. Similar structures have been the object of experimental [17] as well as theoretical [18] studies. In figure 8 the wavelengths of the respective h1e1 transitions are given too. With the exception of the sample with the 20 nm barrier, the photoexcited electrons of the smaller well

[†]This experiment has been performed by W W Rühle from the MPI Stuttgart.



Barriers: 3, 4, 6, 8, 20 nm

Figure 8. GaAs/AlGaAs tunnel structure. Numbers assigned to quantum wells give the wavelengths of their h1e1 transitions.

tunnel non-resonantly into the wider well. For the samples with 3 nm and 4 nm barriers the tunnelling time is much lower than the optical decay time, and hence almost all electrons are tunnelling non-resonantly before optically recombining with holes within the small well. In the sample with 6 nm barriers the tunnelling time approximately equals the optical decay time and for the 8 nm barriers non-resonant tunnelling is less effective than optical recombination. Nevertheless the tunnelling can be detected even in this sample. In the 20 nm barrier sample, however, the tunnelling is completely suppressed. These results are summarized in [16].

We have applied the optical pumping technique to investigate these samples with respect to spin conservation of tunnelling electrons. In an optically detected tunnelling experiment we have measured the transitions due to the quantized states of the smaller well via an increase of the photoluminescence line of the h1e1 transition of the wider well. A spectrum originating from the sample with 3 nm barriers is given by the lower curve of figure 9. We have marked the l1e1 and the h1e1 transitions of the small well in this figure. As we have used circularly polarized light for the excitation we have

found the same resonant transitions in the polarization signal of the luminescence. These transitions exhibit the proper signs, that is positive polarization for the h1e1 transition and negative polarization for the l1e1 transition, as is shown in the upper curve of figure 9. This indicates that the electronic spins are not destroyed during the tunnelling through the 3 nm barrier. By evaluating these experiments quantitatively we obtain the result that the electronic spin is fully conserved during non-resonant tunnelling in the samples with barrier widths of 3 and 4 nm. The spin, however, is fully destroyed in structures with wider barriers. Thus it looks as if a critical barrier thickness, lying between 4 nm and 6 nm, could be defined for spin conservation during non-resonant tunnelling. It is plausible that spin-flip processes occur together with non-resonant tunnelling because the latter is accompanied by multi-stage inelastic scattering processes. It is, however, surprising that the transition between spin conservation and spin destruction occurring with increasing barrier width is such an abrupt one. To clarify whether this phenomenon can be attributed to an increase of the dwell time due to an increase of the barrier width one has to calculate the ratio of the dwell times for 6 nm and 4 nm barriers. For the dwell time we assume a standard formula in which the former exhibits an exponential dependence on the barrier thickness as is given in [19, 20]. As a result of our estimation we obtain that the dwell time at a 6 nm barrier is ten times higher than that at a 4 nm barrier. Our experimental results, however, suggest at least a factor of 100 for this ratio. Hence we conclude that the spin-flip process more likely occurs during the tunnelling through the barrier than during the stay of the photogenerated electrons in the narrow well before tunnelling.

5. Conclusions

Summarizing, we state that spin relaxation processes manifest themselves in the degree of circular polarization of the photoluminescence and that they are very sensitive to the specific conditions of the confinement. Since the interface region is of great importance for spin relaxation in heterostructures we can use optical pumping techniques to characterize this interface quality. The spin lifetime depends more strongly on the interface quality than other observables such as the optical decay time or the photoluminescence linewidth. Optical pumping is very well suited to PLE experiments in confined systems because it allows a clear identification of hh- and lh-related transitions.

Acknowledgment

I wish to thank all those who have contributed to this work: M Kunzer, U Kaufmann, K Köhler, W W Rühle, D J As and W Jantsch.

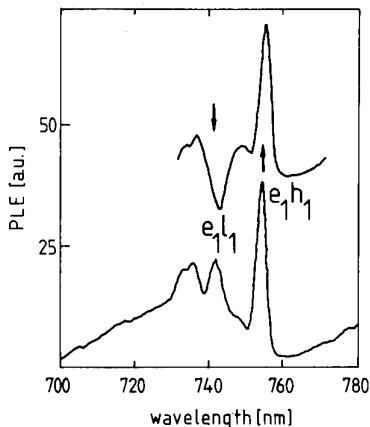


Figure 9. PLE spectra of the narrow well with 3 nm barriers measured via intensity (lower trace) or polarization (upper trace) changes of the luminescence of the wider well. The experiment was done under σ^+ excitation.

References

[1] Herrmann C and Lampel G 1985 *Ann. Phys., Paris* **10** 1117

[2] Cohen-Tannoudji C and Kastler A 1966 *Prog. Opt.* **5** 1

[3] See various chapters in Meier F and Zakharchenya B P (ed) 1984 *Optical Orientation* (Amsterdam: North-Holland)

[4] Miller R C, Kleinman D A, Nordland W A Jr and Gossard A C 1980 *Phys. Rev. B* **22** 863

[5] Miller R C, Gossard A C, Tsang W T and Munteanu O 1982 *Phys. Rev. B* **25** 3871

[6] Miller R C and Kleinman D A 1985 *J. Lumin.* **30** 520

[7] Damen T C, Vina L, Cunningham J E, Shah J and Sham L J 1991 *Phys. Rev. Lett.* **67** 3432

[8] Uenoyama T and Sham L J 1990 *Phys. Rev. Lett.* **64** 3070

[9] Ferreira R and Bastard G 1991 *Phys. Rev. B* **43** 9687

[10] Hendorfer G and Schneider J 1991 *Semicond. Sci. Technol.* **6** 595

[11] Gershoni D, Vandenberg J M, Chu S N G, Temkin H, Tanbun-Ek T and Logan R A 1989 *Phys. Rev. B* **40** 10017

[12] Reithmaier J-P, Höger R, Riechert H, Hiergeist P and Abstreiter G 1990 *Appl. Phys. Lett.* **57** 957

[13] Bastard G 1988 *Wave Mechanics Applied to Semiconductor Heterostructures* (Les Ulis: Les Editions de Physique)

[14] Joyce M J, Johnson M J, Gal M and Usher B F 1988 *Phys. Rev. B* **38** 10978

[15] Arent D J 1990 *Phys. Rev. B* **41** 843

[16] Nido M, Alexander M G, Rühle W W, Schweizer T and Köhler K 1990 *Appl. Phys. Lett.* **56** 355

[17] Leo K, Shah J, Gordon J P, Damen T C, Miller D A B, Tu C W and Cunningham J E 1990 *Phys. Rev. B* **42** 7065

[18] Bavli R and Metiu H 1992 *Phys. Rev. Lett.* **69** 1986

[19] Mendez E E 1987 *Physics and Applications of Quantum Wells and Superlattices* (NATO ASI Series, Series B Physics 170) ed E E Mendez and K von Klitzing (New York: Plenum)

[20] Tada T, Yamaguchi A, Ninomiya T, Uchiki H, Kobayashi T and Yao T 1988 *J. Appl. Phys.* **63** 5491

UHVCVD growth of Si/SiGe heterostructures and their applications

B S Meyerson, K E Ismail, D L Harame, F K LeGoues and
J M C Stork

IBM Research Division, T J Watson Research Center, PO Box 218,
Yorktown Heights, NY 10598, USA

Abstract. The era of integrated circuits based on SiGe heterojunction bipolar transistors arrived with the announcement of a 12-bit digital to analogue converter (DAC) fabricated using an analogue optimization of IBM's SiGe HBT technology. Medium-scale integration was employed, the circuit consisting of approximately 3000 transistors and 2000 passive elements (resistor and capacitors). Operable at 1 GHz, this converter consumes approximately 0.75 W, thus yielding power-delay performance a decade superior to prior devices. It is significant that this DAC was fabricated employing the same technology and toolset as found on a standard silicon-based CMOS product line. In addition to the CMOS toolset, only one unique tool is required to support this technology, a commercial (Leybold-AG) ultrahigh vacuum chemical vapour deposition system for SiGe deposition. It is of interest to note, however, that the processing of these integrated circuits was no different from that employed in fabricating high-performance SiGe high electron mobility transistors (HEMTs), as well as the first N-type SiGe-based resonant tunnelling devices (RTDs), all functional at room temperature. This enables one to combine a wafer-scale manufacturable SiGe-based heterojunction technology with devices that utilize quantum phenomena, made accessible by the use of band offsets and strain-induced band splitting in the Si/SiGe materials system. This new ability to incorporate leading edge developments in SiGe device physics into a standard technology line opens up a host of new areas for exploration.

1. Introduction

It has been long argued that silicon, the mainstay material of modern electronics, has run its technical course. This belief has frequently been supported by the argument that the fundamental limits of the silicon materials system preclude the extension of silicon technology into the future. This view is not without merit, and has in some instances spawned entire fields of endeavour developing the next generation of semiconductors, most notably those based on III-V materials. Being an indirect-gap semiconductor, applications requiring high-efficiency light emission are challenging, although not impossible. It is in fact true that the physical parameters of silicon when employed as an electronic material are well known, and essentially fixed. In searching any text one may find well defined upper bounds for electron mobility, hole mobility and bandgap, all a consequence of the fixed band structure of silicon. To nurture a physicist's/engineer's sense of adventure, as well as preserve a 100 billion dollar investment in silicon-based technology, an added degree of freedom is required so as to step outside the bounds of the silicon materials system. By virtue of the tremendous recent progress [1–5] in working with the silicon:germanium (SiGe) alloy/silicon heterojunction, one can argue that the extra degree of freedom required to drive silicon technology forward has been demonstrated.

Heterojunction applications within technology are well known, based upon the extensive literature in the III-V materials system, notably gallium arsenide and gallium aluminium arsenide (GaAs/GaAlAs). Although SiGe/Si heterojunctions have been studied for over a decade, only recently has the first integrated circuit in a SiGe heterojunction-based technology been reported [1]. The significance of the integrated circuit reported, a 12-bit digital to analogue converter, was not solely its high level of integration or performance. The significance of this result is that the method of fabrication has also been demonstrated to be compatible with resonant tunnelling diodes [6] at the one extreme and conventional CMOS technology on the other, all processed in the toolset utilized in standard silicon technology. This affords the physics community an opportunity to implement devices heretofore viewed as 'exotic' alongside conventional silicon logic—a formidable capability.

In considering the technological applications of novel devices based on a new materials system, it is important to clarify the driving forces behind the many choices that must be made in defining the approach to be taken in generating the new technology. Such choices may be driven by factors as diverse as limitations in film growth capability, optimization of discrete device performance, yield enhancement/defect reduction strategies, circuit performance issues and so on. Rather than focus solely on the outcome of the technical approach taken, it is the

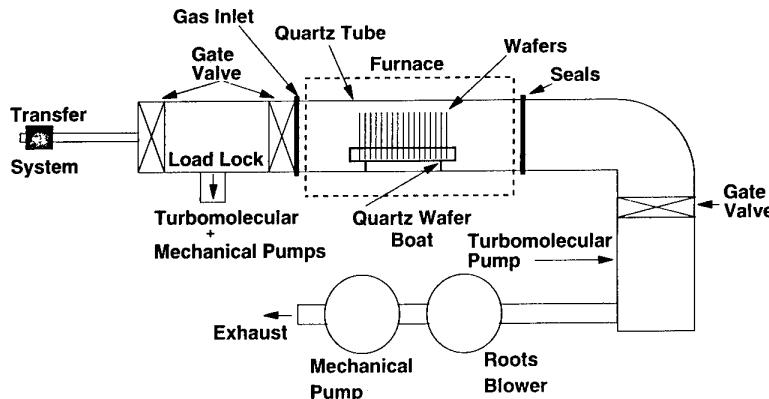


Figure 1. Schematic diagram of a UHVCVD system.

intent here to discuss the fundamental issues that have led us to employ a given device strategy. To provide some background, we will first discuss the film growth method employed in preparing the heterojunction bipolar transistor (HBT) technology described above, including the choices made in the optimization of this technology.

2. Film growth

All the devices discussed in this paper have been fabricated employing the ultrahigh vacuum chemical vapour deposition (UHVCVD) technique [7]. This growth method employs hydrogen passivation [8] to prepare and maintain a high-quality silicon surface prior to the commencement of epitaxy, and employs a chemically specific UHV approach to eliminate contaminants from the actual growth environs. Chemical specificity is important, in that one could go to great lengths to eliminate hydrogen and helium permeation into a growth environment, but such species have no impact on silicon growth chemistry in trace quantities. As such, ultimate vacuum levels of 10^{-9} Torr are adequate, and readily achieved using a quartz vessel as well as simple O-ring technologies, where the residual gas within the growth environment is almost exclusively hydrogen. This simplification dramatically reduces the complexity of the apparatus required in comparison with conventional approaches to UHV. The UHVCVD apparatus is shown schematically in figure 1. Source gases employed are silane, germane, diborane and phosphine. The details of this film growth method have been reviewed extensively elsewhere [9]. In brief, films are deposited in batch mode, employing an isothermal growth environment held in the range of wafer temperatures 400–500 °C, the limits of which correspond to the growth of pure germanium and pure silicon respectively. Being by nature a conformal CVD process, and at no time requiring temperatures in excess of 500 °C, patterned wafers, as required to produce a highly integrated silicon technology, are readily processed by this method. The varied topography and materials found on a silicon device wafer at the point where the base is ideally formed are readily accommodated by the low thermal budget of the UHVCVD growth process.

This enables the attainment of devices having extraordinarily high levels of both performance *and* integration.

3. Silicon:germanium HBTs

Figure 2 shows the reasons for current interest in employing SiGe HBTs in an integrated technology. This trend chart follows the performance gains of discrete, silicon-based bipolar transistors over the past decade, and the limits of performance for SiGe HBT devices lie well above those for conventional silicon devices. This advantage may be extracted in the form of added operational speed in a given circuit application, or low-power operation, where high speeds are maintained even at significantly reduced current levels.

The enhanced performance of the devices represented here stems in large part from the graded SiGe base region, as seen in figure 3, which extends across the base of such a bandgap engineered device. The small total Ge content required in such an approach is the key to attaining high-performance devices without eliminating the possibility of attaining high levels of device integration. Put in perspective, an HBT utilizing a Ge profile ramped as

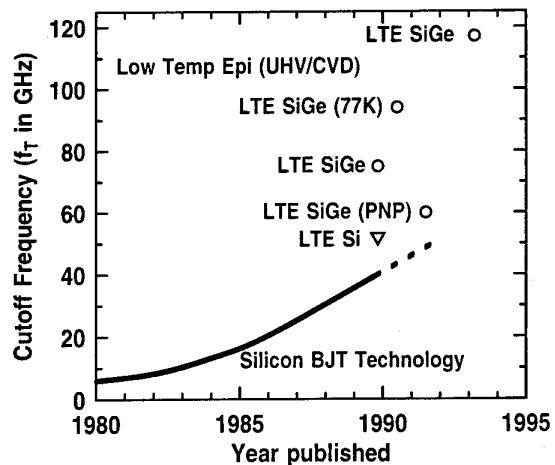


Figure 2. Progress in silicon-based bipolar device speed versus year. A speed differential of a factor of two has been maintained since the discontinuity introduced in 1989 with the advent of high-speed SiGe HBTs.

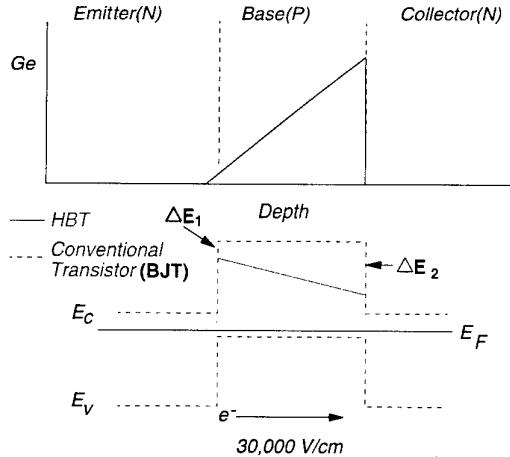


Figure 3. Schematic bandgap diagrams for a graded double heterostructure.

shown over a distance of 300 Å, from 2% at the emitter–base junction to 10% at the peak, results in a built-in electric field of magnitude $30\text{--}40 \text{ kV cm}^{-1}$. This field-assisted transport is employed to reduce the electron base transit time to minimum values. In the extreme of Ge ramps from 0 to 25% over 250 Å, this time has been reduced to subpicosecond levels. The result of that work [10] was non-self-aligned devices having cut-off frequency (f_T) values in the range 110–117 GHz. It is also useful to consider the implications of having two dissimilar heterojunctions, with offsets ΔE_1 and ΔE_2 as independent variables. Although the difference in offsets is employed to generate a field within the base, tailoring the offset at the emitter–base junction enables one to tune the thermal response of such a device in a unique fashion. For example, bandgap shrinkage owing to heavy doping effects in the emitter of a silicon homojunction device cause the gain to decrease exponentially with temperature. The addition of Ge in an amount such that the bandgap of the base is reduced to an identical extent at this point leads to a device whose gain is temperature invariant. This result is of great utility when designing devices for operation in harsh environments, or where a highly stable linear response is required across a wide temperature range. By increasing the germanium content at this junction, inverted thermal behaviour is the norm, where gain rises exponentially below room temperature. The flexibility to achieve all such conditions without a sacrifice in overall HBT performance is a unique attribute of this structure. By contrast, when attempting to produce devices having this level of performance using non-graded Ge profiles within the base of the HBT, twice the Ge dose is routinely employed, and the degree of control demonstrated above is not possible. Furthermore, the consequences of this high Ge dose are significant.

The work of Matthews and Blakeslee [11, 12] provided a standard by which a film under stress may be judged to be unconditionally stable at a given thickness. Above such a ‘critical thickness’ the film is potentially unstable, in that a defect that is formed may

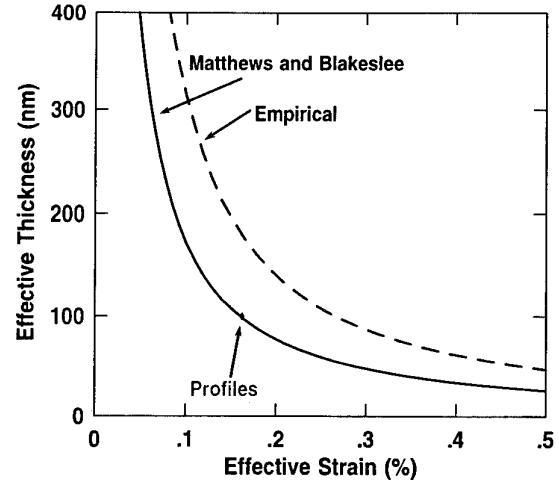


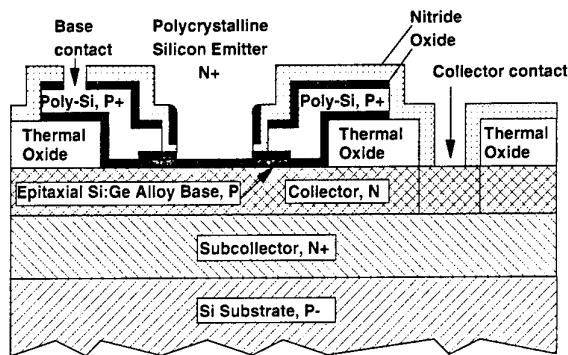
Figure 4. Stability of SiGe alloy layers embedded in an SiGe HBT. Devices which fall on or below the full curve are unconditionally stable.

then move within the lattice. Once nucleated, such a defect may glide within the layer, or spawn others via multiplication, leading to rapid failure of any device or circuit resident in the layer. Computing the stability condition [13] for a commensurate film of SiGe on Si results in the full curve shown in figure 4. The positions of the high-speed devices we have employed in the technology above lie directly on, or below, this stability curve. This is in marked contrast to other attempts to provide this level of device performance in the absence of a graded germanium profile. To achieve similar performance has thus far required the use of a Ge content nearly a factor of two higher, resulting in devices that are inherently unstable. This does *not* mean that such devices are doomed to failure, rather it is a statistical indication of potential for failure.

In addition to the need to determine the optimal Ge profile for use in a SiGe HBT technology, one must choose a structural approach for one’s device design. Two diametrically opposed approaches are shown in figure 5: a mesa device and a planar, integrated structure. Although high levels of performance may be obtained for discrete transistors employing a mesa structure, the complexity of contacting various layers of the device, along with its high topography, prevents its effective use in circuits. By contrast, the great preponderance of high-performance discrete devices, and the only high-performance circuits, have thus far been prepared in planar, self-aligned technologies. It is important to consider the consequences of opting for such a fully integrable bipolar technology.

Although PNP and NPN layers/structures have been deposited by UHVCVD, the uninterrupted sequential deposition of such layers prevents the fabrication of low-topography devices as shown in figure 5. In order to get circuit-level advantage when replacing an ion-implanted base region with one that is to be grown, as in this instance, patterned wafers must be used, with the

Integrated planar bipolar transistor (Graded Bandgap Si:Ge HBT)



Mesa Transistor

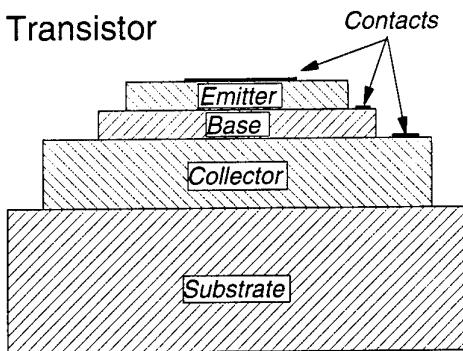


Figure 5. A planar SiGe HBT alongside a more classical mesa device. The chemical complexity of growth over the many exposed materials make planar structures challenging to prepare, but they have proved to be the only devices integrable into circuits.

collector in place and lithographically defined. Similarly, the growth of the emitter directly upon the base is of little value, as this again makes difficult the integration of such devices beyond the single transistor level. It is sometimes stated, in error, that such choices are driven by limitations of growth technique, whereas these choices are in fact a consequence of an overall optimization of the devices and circuits desired.

The issues recited above force one to address the problem of growth on patterned substrates, and are responsible for the emphasis that has been placed on patterned substrate preparation, and thus low-temperature processing. The dopant patterns existent upon such a substrate are prone to redistribution due to excessive thermal cycles, as well as chemical transport due to etching, and this must be guarded against. In contrast to this critical requirement in technological applications, far more straightforward blanket film growth was employed in those instances where unproven physical concepts were tested, such as n-type RTDS, or transferred electron devices. As a final issue, one must choose, from device considerations, whether to employ selective or non-selective film growth.

Employing UHVCVD, it is possible to selectively deposit several hundreds of angstroms of silicon, and/or

several thousands of angstroms of SiGe alloy, under the conditions described above. The choice of selective versus non-selective film growth is thus again driven by device rather than growth considerations. In the instance of raised source drains as in FETs, selective growth of heavily doped films is employed. However, in the instance here, one must deposit uniformly across all surfaces to ensure that the layer comprising the base of the device is continuous into the contact regions. The total base resistance of a device is the sum of all relevant series resistances, and in the optimal case no one term should dominate. The use of ten times the base dopant dose may in fact have no benefit owing to a high base contact or link resistance, yet may degrade performance through scattering or breakdown voltage degradation. A single-film, non-selective approach allows one to achieve very low overall base resistance for a minimum amount of dopant incorporated into the base itself. The selectivity seen in UHVCVD takes the form of an incubation time prior to growth on non-silicon surfaces such as silicon nitride and/or silicon dioxide. This may be defeated by seeding all surfaces with a thin layer of pure silicon grown at relatively low temperature, as was done in this work. It has been observed that growth at higher temperatures, or at high hydrogen dilution, enhances selectivity in this growth method, but this is not exploited here.

Utilizing the aforementioned strategies for film growth and device design, a wide variety of high-performance HBTs, both NPN and PNP, as well as circuits containing them, have been fabricated. Although such devices have been in the literature for roughly five years, circuit results to date, even simple ring oscillators, are notably absent from most work. Substantial work has been devoted to the issues of yield, critical for a viable technology. The leading edge of this effort has produced [14] the results shown in figure 6. The map shown indicates working cells of 30 000 transistors wired with their emitters tied together. A yield of 75–85% is seen for these arrays, a result borne out by the subsequent success of medium-scale integrated circuits in this technology. The devices in these arrays were consistently operable at values of approximately $f_t = f_{max} = 60$ GHz, values well above those available in conventional silicon technology. It was the emphasis on fundamental device design issues, coupled with yield driven Ge profile considerations, that has made this result possible.

Beyond the bipolar devices discussed thus far, much progress has been made in more ‘exotic’ devices, such as silicon channel HEMTs. In spite of the advanced and somewhat speculative nature of this new area in SiGe-based devices, work is proceeding with the goal of integrating additional function alongside or in place of the HBT work to date. It is on this most recent aspect of this field that the remainder of this paper will focus.

4. Exploratory SiGe devices and structures

Employing film growth and processing methods as utilized in bipolar technology as above, recent more

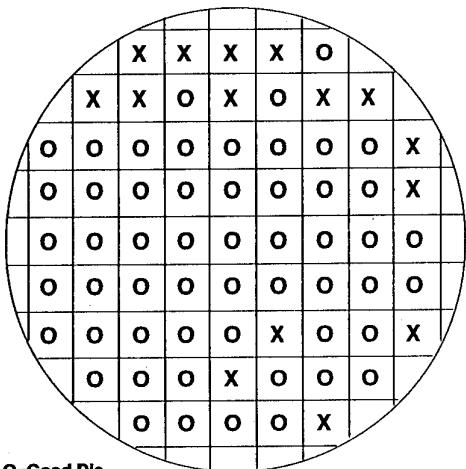


Figure 6. A yield map for 30 000 transistor arrays wired as common emitter devices. Note that some test sites which 'failed' were at the wafer edge, and in some cases fell outside of the edge of the wafer. Without screening out such points, yield reached 75%. Good and failed dies are indicated by circles (○) and crosses respectively.

exploratory work has focused on the implementation of strain-induced bandgap engineered devices. In the most direct application of such work, Ge channel PMOS devices were fabricated, utilizing the bandgap reduction and valence band offset induced when SiGe is deposited commensurate to silicon. Such devices are based upon hole confinement, and by modulation doping the device, again as is done in the III-V system, significant performance gains were possible. Similarly, using the strain-induced band splitting in the SiGe/Si heterojunction, combined with new methods for defect suppression by compositional grading [15–17], high-mobility two-dimensional electron gas structures have been fabricated. The recent work of Ismail *et al* [3] has focused on the room-temperature optimization of such transport, for electrons and holes. The result of this work has been fully confined room-temperature electron and hole gas mobility values several times greater than the upper bound for intrinsic silicon; for example 2700–3000 cm² V⁻¹ s⁻¹ mobility for electrons, at room temperature, in a tensile silicon channel on a relaxed SiGe substrate. Similarly, values in excess of 800 cm² V⁻¹ s⁻¹ [18] were produced in a compressively strained 80% SiGe layer atop the same structure. The challenge for this work will be the implementation of both such structures in a manner that allows them to be merged into a workable HEMT-like CMOS technology.

In working with these structures, both electron and hole transport showed clear evidence of the spontaneous oscillations associated with transferred electron devices. In vertical transport through a compressive layer of 20% SiGe, shown in figure 7, and in lateral transport in a compressive 80% Ge layer (a Corbino ring structure),

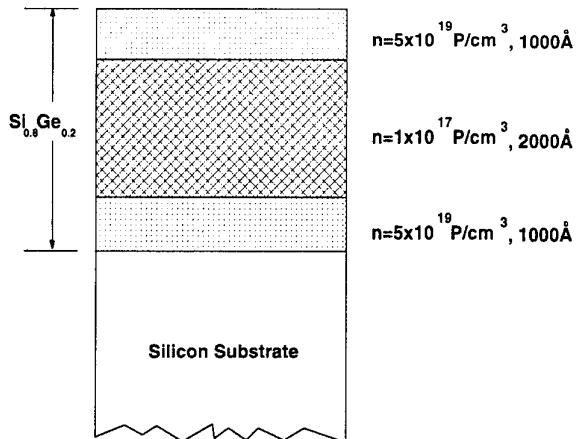


Figure 7. A cross section of a simple transferred electron device oscillator, employing compressive strain to lift the conduction band degeneracy in the SiGe alloy. These preliminary results have been reproduced at higher strain levels in planar, compressive, 80% SiGe channels. The SiGe alloy is commensurate to underlying silicon.

spontaneous oscillations were observed [19] for applied fields in the range of 50 kV cm⁻¹ for electrons and 2 kV cm⁻¹ for holes. In the instance of the Corbino structures, the peak to valley ratio in the oscillators was greater than 10:1. This first evidence of the phenomenon is consistent with expected splitting of light and heavy states for both electrons [20] and holes in such strained layers. It must be remembered that the high-mobility structures now common in the literature stem from the strain-induced splitting of light- and heavy-mass conduction bands, in the plane of a tensile silicon layer. The same splitting caused by elongation along the z axis in a constrained layer lifts this degeneracy, and allows for the interlevel scattering, for example negative differential conductance, which is required to produce an oscillator. In that both structures were not embedded in an on-wafer RC circuit, the probe station employed limited any frequency performance data to 500 MHz. The structural simplicity of these structures offers considerable fabrication advantage over present oscillators such as IMPATT diodes, but their preliminary nature prevents any definitive assessment of the relative merits of these two devices. Being the most recent in the series of novel devices based on bandgap engineering, it is premature to speculate on their implementation within a specific technology, but the utility of an embedded, simple, on-chip oscillator is beyond doubt.

5. Summary

The issues being addressed in the transformation of SiGe-based heterojunction devices from research topic to a viable technology have been explored. It is noted that this process is actually under way, but by no means complete. The technical issues driving the approach taken to commercial implementation are based upon

fundamental considerations of device function, yield and optimization, a considerable step forward from prior restrictions owing to limitations associated with film preparation. Of equal import, the same strategies employed in reducing SiGe HBTs to practice are being further exploited in developing novel devices, such as resonant tunnelling diodes and the oscillators reported herein.

References

- [1] Harame D L *et al* 1993 *IEEE IEDM Tech. Digest, Proc. IEDM 93* pp 71–4
- [2] Ismail K, Meyerson B S and Wang P J 1991 *Appl. Phys. Lett.* **58** 2117–19
- [3] Ismail K E, Nelson S F, Chu J O and Meyerson B S 1993 *Appl. Phys. Lett.* **63** 660–2
- [4] Schaffler R, Tobben D, Herzog H, Abstreiter G and Hollander B 1992 *Semicond. Sci. Technol.* **7** 260–6
- [5] Konig U and Schaffler F 1991 *Electron. Lett.* **27** 1405–6
- [6] Ismail K, Meyerson B S and Wang P J 1991 *Appl. Phys. Lett.* **59** 973–5
- [7] Meyerson B S 1986 *Appl. Phys. Lett.* **48** 797–9
- [8] Meyerson B S, Himpel F J and Uram K J 1990 *Appl. Phys. Lett.* **57** 1034
- [9] Meyerson B S 1992 *Proc. IEEE* **80** 1592–608
- [10] Crabbe E *et al* 1993 *113-GHz Ft graded-base SiGe HBTs* (Santa Barbara, CA: TMS)
- [11] Matthews J and Blakeslee A E 1974 *J. Crystal Growth* **27** 118–25
- [12] Matthews J and Blakeslee A E 1975 *J. Crystal Growth* **32** 265–73
- [13] Stiffler S R, Comfort J H, Stanis C L, Harame D L, deFresart E and Meyerson B S 1991 *J. Appl. Phys.* **70** 1416
- [14] Harame D L *et al* 1992 *IEEE IEDM Tech. Digest, Proc. IEDM 86* (Washington, DC: IEEE) pp 19–22
- [15] Meyerson B S, Uram K J and LeGoues F K 1988 *Appl. Phys. Lett.* **53** 2555–7
- [16] Koenig U, Boers A J, Schaffler F and Kasper E 1992 *Electron. Lett.* **28** 160–2
- [17] Xie Y H, Monroe D, Fitzgerald E A, Silverman P J, Thiel F A and Watson G P 1993 *Appl. Phys. Lett.* **63** 2263
- [18] Ismail K E, Chu J O and Meyerson B S 1994 *Appl. Phys. Lett.* **64** 3124–6
- [19] Ismail K E, Meyerson B S and Chu J O 1994 to be published
- [20] Abstreiter G, Brugger H, Wolff T, Jorke H and Herzog H 1985 *Phys. Rev. Lett.* **54** 2441–3

Photo- and electroluminescence in short-period Si/Ge superlattice structures

Janos Olajos[†], Jesper Engvall[†], Hermann G Grimmeiss[†], Ulrich Menczigar^{‡\$}, Markus Gail[‡], Gerhard Abstreiter[‡], Horst Kibbel^{||}, Erich Kasper^{||} and Hartmut Presting^{||}

[†] Department of Solid State Physics, Lund University, Box 118, S-221 00 Lund, Sweden

[‡] Walter Schottky Institut, TU-München, Am Coulombwall, D-85748 Garching, Germany

^{||} Daimler-Benz Research Center, D-7900 Ulm, Germany

Abstract. Interband optical transitions have been studied in a variety of short-period Si/Ge superlattice structures by means of photocurrent spectroscopy, infrared absorption, photo- and electroluminescence. Furthermore, the bandgap photoluminescence from strain-adjusted Si_mGe_n ($m = 9, 6, 3$; $n = 6, 4, 2$) superlattices was studied under applied hydrostatic pressure. The strain adjustment was achieved by a thick, step-graded $\text{Si}_{1-x}\text{Ge}_x$ buffer layer resulting in an improved quality of the superlattice with respect to dislocation density. The hydrostatic pressure dependence was modelled using an approach based on deformation potentials and effective-mass theory. In samples annealed at 500 °C and higher, a systematic shift of the bandgap was observed which is discussed in terms of a process involving interdiffusion of the Si and Ge atoms.

Bandgap-related electroluminescence was observed in mesa diodes at room temperature, whereas the photoluminescence disappeared at about 40 K. The electroluminescence from samples based on different buffer-layer concepts is compared.

Apart from the strain-symmetrized Si/Ge superlattices, another structure that has been proposed to act as an efficient, light-emitting device in the Si-based systems is an ultrathin Ge layer (1–2 monolayers) embedded in bulk Si. We report on the electroluminescence spectra at various temperatures from a sample based on this concept, namely a layer sequence consisting of two periods of $\text{Si}_{17}\text{Ge}_2$ grown pseudomorphically on an n^+ Si substrate. A very intensive, well resolved electroluminescence was obtained at 55 K from the ow.

1. Introduction

The system of Si/Ge and $\text{Si}/\text{Si}_{1-x}\text{Ge}_x$ strained-layer superlattices (SLSS) and quantum wells (QWS) offers exciting perspectives from both the basic research and the technological points of view. The large lattice mismatch of about 4% between Si and Ge to some extent sets a limit to possible structures, but simultaneously it offers an effective tool for bandgap engineering since the band structure of the novel structures is strongly influenced not only by the quantum mechanical confinement but also the actual strain situation. The growth of lattice-mismatched materials results either in a strained-layer configuration, if the grown layer is thinner than a critical thickness, or in a strain relaxation by the formation of misfit dislocations. Consequently, the realization of a Si/Ge-based structure for optical or electrical applications must involve either a very thin epitaxial layer which does not exceed the critical thickness or, if a thicker layer is needed, a method which during

growth reduces the density of threading dislocations that extends into the active region. An additional difficulty that arises when truly short-period structures are considered is the intermixing of the two atomic species.

In this paper we will discuss recent progress concerning the understanding of the optical and electro-optical properties of Si/Ge SLSS based on the two different approaches described above.

2. Strain-symmetrized Si_mGe_n superlattices

One of the driving forces behind the rather extensive experimental [1–5] and theoretical [6–14] work done on these structures is the suggestion, based on rather simple ‘Brillouin-zone folding’ arguments, that a ‘quasi-direct’ semiconductor material could be obtained from originally indirect semiconductors if they were grown in a very special way, involving alternating layers of the two different species [6]. In the case of Si/Ge, the basic idea is to grow a superlattice with a period of 10 monolayers (MLs). For this periodicity it is assumed that

^{\$} Present address: Institut für Halbleiterphysik, Frankfurt/Oder, Walter Korsing Strasse 2D-15232 Frankfurt (Oder), Germany.

the conduction band (CB) minima, which are located at 0.85 in the Γ -X direction for bulk Si, are folded back to the Γ point. In this simple model, an enhanced oscillator strength is expected for these so-called ‘quasi-direct’ transitions between the valence band and the two-fold degenerate conduction band edge. A large number of more elaborate band structure calculations [7, 10–14] also predict an enhanced oscillator strength for the transitions. The resulting band structure of a short-period, strained-layer superlattice will be determined, however, not only by zone folding and confinement but also, very importantly, by the actual strain situation. For the case of Si/Ge SLSS grown lattice-matched to Si [001] substrates, it is found that the in-plane (x, y) four-fold conduction band minima are lowest in energy. For these minima, no zone folding is expected. In order to bring the two-fold (z) CB minima below the four-fold in-plane CB minima, deformation potential theory together with an effective-mass (EM) calculation gives that the Si layers in the SLS must be under a lateral tensile strain. This can be accomplished if the SLS is grown on a substrate with a lattice constant in between that of Si and Ge. A substrate material, other than Si in Si-based microelectronics is, however, excluded for obvious reasons. Strain-adjustment on a Si substrate can still be achieved by the concept of ‘virtual substrates’ [15] which involves a partly relaxed $\text{Si}_{1-x}\text{Ge}_x$ alloy buffer layer. The Ge concentration in the buffer can, furthermore, be chosen so that the resulting lattice constant of the alloy provides a strain distribution for the SLS, which is grown on top of the buffer, such that the lateral strains in the Si and Ge layers are of equal magnitude but of opposite sign. In this way the integral of the built-in elastic energy in the SLS is zero and no further relaxation occurs, even for very thick superlattices. The relaxed $\text{Si}_{1-x}\text{Ge}_x$ alloy buffer layer will, however, contain dislocations to a high density which will extend into the SLS region and reduce the performance of the superlattice. The quality of the buffer layer is governed by elastic-strain-driven phenomena such as dislocation generation and surface waviness [16]. It has been shown during the last years that thick buffers with a step-wise or linearly increasing Ge concentration have an improved crystalline quality and a threading dislocation density almost four orders of magnitude lower than a thin buffer layer. A further improvement of the structural properties of the SLS, which reduces the intermixing of the Si and Ge atoms at the interfaces in SLSS and preserves the two-dimensional growth mode in alloys, can be obtained by using surfactants. The segregation of Ge atoms can be efficiently suppressed by deposition of only 1 ML of Sb [17]. Employing the surfactant technique, the Si/Ge SLSS can be grown at a much higher temperature, around 500 °C, without any significant interdiffusion. This is important from the point of view of the luminescence properties, since it has been observed that efficient PL and MBE-grown structures can usually be obtained from samples grown at 500 °C or higher. We will in the next sections discuss the optical properties of a set of SLSS which were grown under these optimized conditions.

3. Experiment

The samples used in this work were grown in a Si MBE chamber at Daimler-Benz Research Center and comprised both undoped Si_mGe_n SLSS and superlattices grown in the depletion region of p-n junctions. For the undoped SLSS, a series was grown with m/n ratios of nominally 9/6, 6/4 and 3/2. The buffer layer consisted of a step-graded alloy buffer 6500 Å thick, followed by a 5000 Å thick $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy. In the step-graded region of the buffer the Ge content was step-wise increased (3%/500 Å) while the temperature was continuously lowered from 600 °C to 520 °C. Prior to the growth of the SLS 1 ML of Sb was deposited to act as a surfactant. The dislocation density, measured with cross-sectional transmission electron microscopy (TEM) was reduced by at least three orders of magnitude compared with previously studied Si_mGe_n SLSS [3] grown on partly relaxed alloy buffer layers. The deviation of all samples from nominal strain and period lengths was less than 10%. The samples were characterized in detail using x-ray diffraction (XRD), TEM and Raman scattering. The structural values obtained from these measurements have been reported elsewhere [18]. In all samples a slight interdiffusion between the Si and Ge layers is present due to the relatively high growth temperatures of the superlattice layers (520 °C). This is particularly important for the sample with the shortest superlattice period (Si_3Ge_2). This sample should be best considered as an $\text{Si}_{1-x}\text{Ge}_x$ alloy with a strong modulation of the concentration. The PL was excited with the 457 nm or 556 nm line of an Ar⁺ laser. The samples were measured at various temperatures in a variable-temperature liquid He cryostat. The luminescence and short-circuit current spectra were measured with a variety of grating monochromators and a BOMEM DA8 Fourier transform spectrometer (FTIR).

4. Results and discussion

Figure 1 shows the PL spectra for three SLSS investigated, namely $(\text{Si})_9/(\text{Ge})_6$, $(\text{Si})_6/(\text{Ge})_4$, $(\text{Si})_3/(\text{Ge})_2$ and a reference $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy grown under the same conditions as the superlattices. The dominant luminescence signals of each sample consist of a no-phonon (NP) peak at higher energies and a transverse-optical (TO) phonon replica. The spectra are corrected with regard to the response of the spectroscopic system. It is interesting to note that the superlattice with the 10-ML periodicity, the $(\text{Si})_6/(\text{Ge})_4$, exhibits the strongest PL for the NP line as compared with the TO replica. The PL properties of these samples have been discussed in detail in an earlier communication [5] showing that the luminescence was consistent with a model involving near-bandgap recombination over excitons localized to potential fluctuations caused by variations in the strain and Ge concentration. Taking into account the fact that the thickness of the SLS region is much smaller than that of the $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy layer the intensity of the LE_{NP} line is enhanced by about

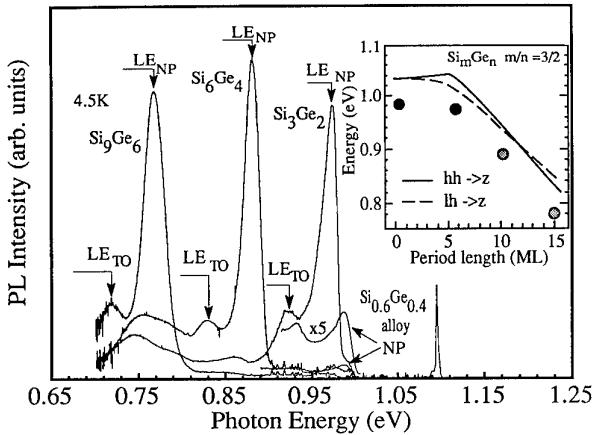


Figure 1. PL spectra for Si_9Ge_6 , Si_6Ge_4 , Si_3Ge_2 and an $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy.

a factor of 150. At lower energies defect-related features can be observed in all three samples. The energy gaps between the hh (heavy-hole) and lh (light-hole) valence band states and the two-fold degenerate (z) conduction band states for strain-symmetrized Si_mGe_n SLSS with a nominal value of $m/n = 1.5$ as calculated with the EM approach are shown in the inset of figure 1. The systematic shift of the bandgap PL with increasing period length for the Si_9Ge_6 and Si_6Ge_4 samples is well described by the calculation. Similar systematic shifts of the onset of interband transitions on both undoped superlattices [19] and diode samples [4] have been studied earlier. Since the PL is observed very close to these onsets it is possible to attribute the luminescence as (near-)bandgap luminescence. To give further strength to this assumption and to learn more about the physics behind the observed spectra additional experiments involving hydrostatic pressure and isochronal annealing were done.

The application of an external hydrostatic pressure to a semiconductor results in a change in the lattice constant and, therefore, in the bond lengths, which influences the band structure and electrical properties of the solid. For example, energy shift of a luminescence line occurs as a function of applied hydrostatic pressure in different ways depending on whether it originates from a (near-)bandgap transition or a recombination over a deep defect level. A detailed study of the photoluminescence spectra of these samples under applied hydrostatic pressure was recently carried out [18]. An example of such a measurement on the nominal Si_6Ge_4 is shown in figure 2. It is seen that the NP lines shift linearly to *lower* energies without any significant change in the amplitude. The PL spectra exhibited, in addition to the NP lines, other somewhat broader features, which at atmospheric pressure were weakly observed at lower energies. These lines gained in intensity with increasing pressure and showed a *positive* pressure coefficient, suggesting that other recombination processes, i.e. *non-interband* transitions are involved in these features. The chemical origin of such defect-related transitions is

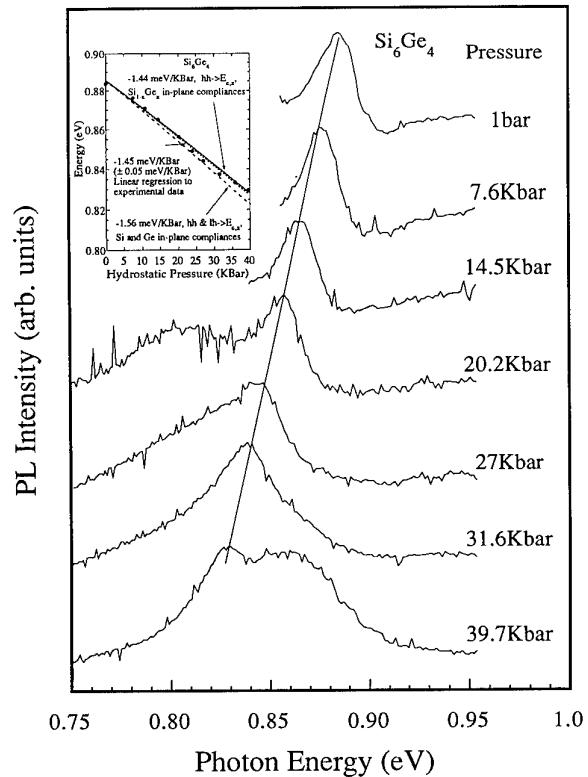


Figure 2. PL spectra as a function of applied hydrostatic pressure for the Si_6Ge_4 sample. In the inset comparison between the calculated pressure dependences and the experimental energy shift of the NP line is shown.

at present unknown. However, these experiments have made it possible to show for the first time experimentally that *both* defect-related PL and bandgap luminescence are observed in short-period Si/Ge superlattice systems. The stress dependence was modelled with the methods described in the previous sections. An externally induced, pure hydrostatic strain does not split the conduction and valence band edges; however, a shift of the weighted averages is introduced. The splittings from the weighted averages due to the built-in biaxial compression (Ge layers) and tension (Si layers) as well as to the slightly non-hydrostatic part of the external pressure have to be calculated for the valence-band multiplet at Γ_{25} , and for the conduction-bands at both L and Δ^1 in the Brillouin zone. The hydrostatic influence is then introduced as a *relative* shift of the valence- and conduction-band weighted averages since the *absolute* quantities are very difficult to determine experimentally as well as theoretically. The differences in band offsets due to the extra hydrostatic pressure is then included around the weighted averages of the valence and conduction bands. The calculated pressure dependences for the Si_6Ge_4 superlattice and the experimentally observed energy shifts of the NP lines are compared in the inset of figure 2. It is readily seen that the parameter-free theoretical slopes obtained from an approach where the Si and Ge layers in the SLS are treated as having the in-plane compliance of the thick $\text{Si}_{1-x}\text{Ge}_x$ alloy buffer (full line) coincide

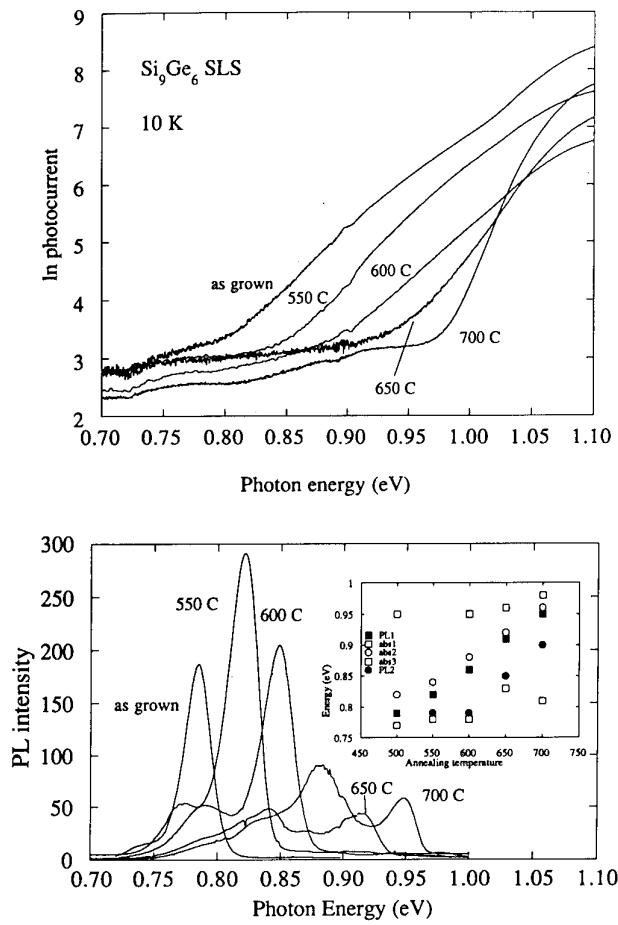


Figure 3. Photocurrent spectra (top) and PL spectra (bottom) for a Si_9Ge_6 SLS as a function of isochronal annealing at various temperatures. In the inset in the lower part the onset energies as determined from a fitting procedure from the photocurrent spectra are compared with various PL energies.

markedly well with a linear regression of the experimental data (dotted line). For comparison, the data from the model where the Si and Ge layers are treated as free-standing layers are also shown (broken line). In this case the agreement between theory and experiment is less favourable.

A further experimental study of the electronic properties was performed by isochronal annealing. The samples were annealed for 1 h at 550, 600, 650, 700 and 750 °C. The PL spectra together with the spectral dependence of the absorption coefficient as measured with the parallel photoconductivity method for the Si_9Ge_6 SLSs are shown in figure 3. The annealing results in a shift of the bandgap-related PL lines (NP and TO) as well as of the onset of interband absorption to higher energies. At higher annealing temperatures, broad PL bands are observed below the NP line. In a simple interdiffusion model [20] it is assumed at an initial stage of annealing that the Si atoms are diffusing into the Ge layers more easily than the Ge atoms into the Si layer. This results in an $\text{Si}/\text{Si}_{1-x}\text{Ge}_x$ superlattice and accordingly in a higher bandgap due to the lower band offsets.

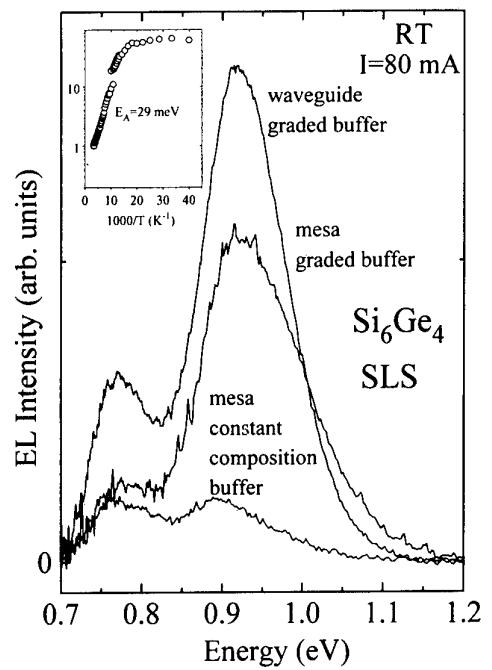


Figure 4. A comparison of the EL intensity from Si/Ge SLSs grown under different conditions.

The experimental bandgaps in the differently annealed samples can easily be determined from a fitting procedure to the absorption spectra. These energies (open symbols) are compared with the various PL transitions (full symbols) in the inset of figure 3. It is seen that at the final stage of annealing additional strain relief occurs in addition to the final interdiffusion. This can be observed both from the appearance of the D_1 and D_2 dislocation-related PL lines and from the slight shift of the bandgap of the underlying $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy to higher energies (the onset with highest energy).

Recently, enhanced electroluminescence at room temperature from a Si_6Ge_4 strain-adjusted superlattice p^+n junction in the region 1.3–1.7 μm (0.70–0.95 eV) has been observed [21]. Absorbance measurements, capacitance–voltage ($C-V$), intensity versus current ($I-V$) and voltage–intensity ($V-L$) measurements supported the assignment of the observed EL (electroluminescence) to originate both from transitions over a defect-related recombination centre (0.77 eV) and from band-to-band recombination processes (0.88 eV) in the SLS. In figure 4, room-temperature electroluminescence spectra from recent samples, based on different buffer layer designs, are shown. For mesa diodes at normal incidence it is seen that a considerable increase in the bandgap emission at about 0.9 eV is obtained from samples grown on the step-graded buffer for a constant current. This is interpreted as a lower concentration of non-radiative recombination channels due to defects for these samples. Further increase in the light output is obtained if the superlattice is overgrown with a thick $\text{Si}_{0.6}\text{Ge}_{0.4}$ alloy waveguide layer and fabricated into a waveguide edge emitter. The bandgap EL intensity is temperature depend-

ent. It decreases exponentially at higher temperatures with an activation energy of about 29 meV.

5. Thin Ge layers embedded in Si

Another approach to increase the oscillator strength in Si/Ge microstructures is based on the observation that an originally forbidden zero-phonon transition in bulk indirect semiconductors can be made allowed if there is some mixing of bulk momenta, as occurs for example when an electron is scattered at an Si–Ge interface [9]. Consequently, it may be possible to produce such transitions simply by growing one or more thin Ge layers within a thick Si matrix. Since it is mainly the Si–Ge interface which is of importance, the total thickness of the Ge layer can be kept low (2–3 ML), and thus the critical thickness does not have to be exceeded. Also strong confinement for the holes is predicted for these structures.

5.1. Results and discussion

Layer sequences based on this concept were grown at the Daimler-Benz Research Center and consisted of a series of $\text{Si}_{17}\text{Ge}_2$ structure repeated two, five and ten times in the depletion region of p-i-n diodes. The electroluminescence spectra of this series taken at 56 K with an injection current density of 20 A cm^{-2} is shown in figure 5. A strong EL feature is observed for all samples which consists of a doublet peak at around 1.0 eV. The peak position is in good agreement with the onset of interband transitions in the SLS, and therefore we believe that the luminescence is a (near-) bandgap emission.

In the inset of figure 5, a comparison is shown between the EL spectrum of the sample with two periods and the PL from an ordinary pseudomorphic $\text{Si}_{1-x}\text{Ge}_x$ alloy quantum well. PL for such structures has been studied extensively over the last years and has been interpreted as consisting of a no-phonon line and a set of phonon replicas reflecting the various possible combinations of Si–Si, Si–Ge and Ge–Ge vibrations in a $\text{Si}_{1-x}\text{Ge}_x$ alloy with fairly low Ge concentration. In the inset it is clearly seen that energy difference between the two peaks in the EL spectra from the $\text{Si}_{17}\text{Ge}_2$ SLS corresponds to the difference between the no-phonon line and the TO replica of the Si–Ge vibration in the alloy quantum well. It is, therefore, possible to attribute the EL spectrum of the $\text{Si}_{17}\text{Ge}_2$ structure as consisting of a no-phonon line and a TO phonon replica from Si–Ge vibration from atoms at the interface of the Ge layers. We find that only the Si–Ge mode is observed, which is essentially the only bond present at such an interface. This also gives further strength to the theoretical prediction about the importance of the Si–Ge interface. With increasing number of periods, the EL signal shifts to lower energies and decreases in intensity. The intensity decrease could be understood in terms of lower carrier concentration in the thicker samples while the shift in energy may be due to

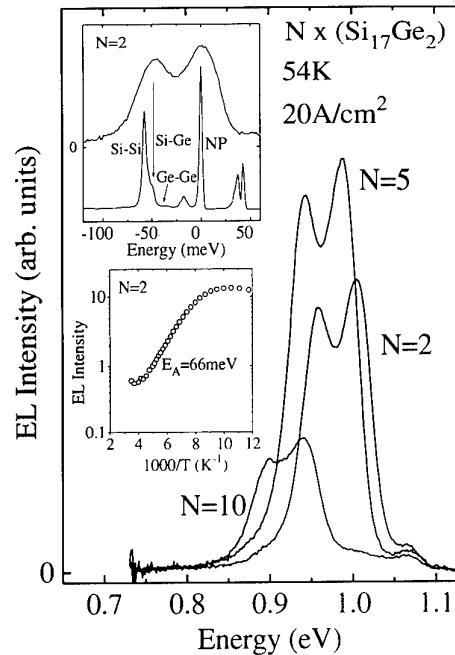


Figure 5. EL spectra for three Si/Ge MQWs. In the insets the temperature dependence of the EL as well as a PL spectrum from a $\text{Si}_{1-x}\text{Ge}_x$ alloy QW is shown.

interacting wells; however, further investigations are necessary before definite assignments can be made. The EL from these structures is strongly dependent on temperature at both lower and higher temperatures. A decrease of the SLS electroluminescence is found at higher temperatures, as can be seen in the inset of figure 5. The intensity drops exponentially with an activation energy of about 66 meV, which is considerably larger than for the superlattices discussed in the previous section.

Acknowledgments

The authors acknowledge valuable discussions with M Jaros and C Tserbak. This work was partly financed by the European Basic Research programme (ESPRIT) under contract no P7128 and by the Swedish National Science Council, the Swedish Board for Technical Development, as well as the Bank of Sweden Tercentenary Foundation.

References

- [1] Pearsall T P, Bevk J, Feldman L C, Bonar J M, Mannaerts J P and Ourmazd A 1987 *Phys. Rev. Lett.* **58** 729
- [2] Pearsall T, Vandenberg J M, Hull R and Bonar J M 1989 *Phys. Rev. Lett.* **63** 2104
- [3] Zachai R, Eberl K, Abstreiter G, Kasper E and Kibbel H 1990 *Phys. Rev. Lett.* **64** 1055
- [4] Olajos J, Engvall J, Grimmeiss H G, Menczigar U, Abstreiter G, Kibbel H, Kasper E and Presting H 1992 *Phys. Rev. B* **46** 12857

- [5] Menczigar U, Abstreiter G, Olajos J and Grimmeiss H G 1993 *Phys. Rev. B* **47** 4099
- [6] Gnutzmann U and Clausecker K 1974 *Appl. Phys.* **3** 9
- [7] Turton R J, Jaros M and Morrison I 1988 *Phys. Rev. B* **38** 8397
- [8] Jaros M, Wong K B and Turton R J 1990 *J. Electron. Mater.* **19** 35
- [9] Turton R J and Jaros M 1990 *Mater. Sci. Eng. B* **7** 37
- [10] Schmid U, Christensen N E, Cardona M and Alouani M 1990 *Proc. 20th Int. Conf. on the Physics of Semiconductors* ed E M Anastassakis and J D Joannopoulos (Singapore: World Scientific) p 865
- [11] Schmid U, Christensen N E and Cardona M 1990 *Phys. Rev. B* **41** 5919
- [12] Schmid U, Lukes F, Christensen N E, Alouani M, Cardona M, Kasper E, Kibbel H and Presting H 1990 *Phys. Rev. Lett.* **65** 1933
- [13] Schmid U, Christensen N E, Alouani M and Cardona M 1991 *Phys. Rev. B* **43** 14597
- [14] Tserbak C, Polatoglou H M and Theodorou G 1993 *Phys. Rev. B* **47** 7104
- [15] Kasper E 1986 *Surf. Sci.* **174** 630
- [16] Heigl G, Span G and Kasper E 1992 *Thin Solid Films* **222** 184
- [17] Fujita K, Fukatsu S, Yaguchi H, Igarashi T, Shiraki Y and Itoh R 1991 *Mater. Res. Soc. Symp. Proc.* **220** 193
- [18] Olajos J, Yia Ying-Bo, Engvall J, Grimmeiss H G, Kibbel H, Presting H and Kasper E 1994 *Phys. Rev. B* **49** 2615
- [19] Olajos J, Engvall J, Grimmeiss H G, Kibbel H, Kasper E and Presting H 1992 *Thin Solid Films* **222** 243
- [20] Schorer R, Friess E, Eberl K and Abstreiter G 1991 *Phys. Rev. B* **44** 1772
- [21] Engvall J, Olajos J, Grimmeiss H G, Presting H, Kibbel H and Kasper E 1993 *Appl. Phys. Lett.* **63** 491

Investigation of luminescence in strained SiGe/Si modulated quantum well and wire structures

Yasuhiro Shiraki and Susumu Fukatsu

Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153, Japan

Abstract. Luminescence properties associated with strain-induced band modification in SiGe/Si heterostructures such as quantum wells (qws) and quantum wires (owrs) are reviewed. Among several issues concerning formation of highly luminescent SiGe materials, surface segregation is shown to be the main cause of deteriorating interface integrity. To resolve the problem, a new technique called segregant-assisted growth (SAG) is proposed. SAG and gas-source MBE (GSMBE), which is considered to be quasi-SAG, are shown to provide high-quality SiGe/Si heterostructures with abrupt interfaces. Highly efficient band edge emissions are observed in not only type-1 but also spatially indirect type-2 qws. The energy shift in qws is discussed based upon the band modification due to surface segregation during growth. The ratio between the no-phonon (NP) peak and its phonon replica (r0) in the edge emission reflects the nature of the qws formed with alloy materials. The coupling of qws and evolution of superlattices are well understood based on the effective-mass approximation by precisely taking into account the band alignment. qwrs are well fabricated on V-groove patterned substrates and luminescent properties very different from those of qws are observed. These findings indicate the high potential of SiGe/Si heterostructures, not only in scientific areas but also in device applications.

1. Introduction

SiGe/Si heterostructures are attracting much attention not only because they have great potential for device applications but also because they open up a new field of semiconductor physics where the strain plays an important role in determining the band structure. The bandgap as well as the band alignment changes depending on the strain distribution in the heterostructures. This can be well studied from the luminescence measurements of qws if their crystal quality is high enough to provide band edge emissions. Therefore, it is natural that there is a strong demand for realization of high-quality strained SiGe/Si quantum wells (Qws) and superlattices (SLS). In this review work done for this purpose at the Research Center for Advanced Science and Technology, The University of Tokyo, is described.

There are several important issues concerning the formation of highly luminescent SiGe materials. This first one is the critical thickness of the strained layers, that is, the maximum thickness where coherent epitaxial growth can take place. Recently, formation of abrupt heterointerfaces has also been recognized to be a critical issue, and surface segregation is known to be the main cause of the interfacial smearing during molecular beam epitaxial

(MBE) growth. Surface segregation is one of the surface reactions between impinging atoms and surface atoms of substrates. This phenomenon was first recognized to be important when MBE layers were doped with some kinds of impurities [1, 2]. It should be pointed out, however, that it is more or less generally observed not only for doping but also for heterointerface formation. To suppress the segregation, low-temperature growth was thought to be essential. However, the low growth temperature sacrifices the crystal quality and luminescent material can hardly be obtained. There were several attempts to improve the crystal quality of MBE-grown materials, including post-growth annealing. Interestingly, the annealing resulted in the appearance of a broad emission peak which was not observed in SiGe alloy bulk crystals, and band edge emissions, which were confirmed in SiGe epitaxial films formed by such gaseous growth methods as rapid thermal chemical vapour deposition [3], were never observed [4]. Recently, several groups including ours have demonstrated that if we employ a high growth temperature, far beyond the commonly accepted growth temperatures, highly luminescent SiGe/Si Qws providing edge emissions can be formed [5–8]. Growth conditions for luminescent SiGe/Si Qws will therefore be discussed first.

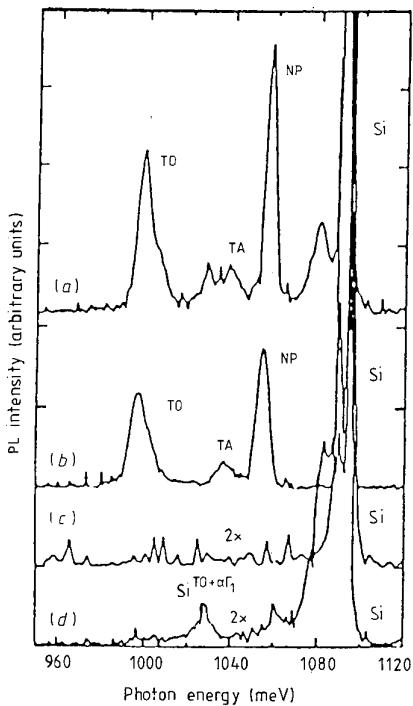


Figure 1. Growth temperature (T_s) dependence of PL spectra of $\text{Si}_{0.80}\text{Ge}_{0.20}/\text{Si}$ QWS formed by solid-source MBE. $T_s = 700^\circ\text{C}$ (a), 660°C (b), 580°C (c) and 500°C (d).

2. Growth conditions for luminescent SiGe/Si QWS and surface segregation

Figure 1 shows photoluminescence (PL) spectra of SiGe/Si QWS grown by solid-source MBE at various temperatures [5]. It is clearly seen in this figure that the high growth temperature gives rise to intense bandgap luminescence; that is, NP and TO peaks, which are respectively attributed to no-phonon and transverse-optic phonon-assisted emissions originating from SiGe well layers [9], are well resolved in the case of 700°C growth. On the other hand, no band edge luminescence is observed in samples grown at temperatures lower than 600°C . These high growth temperatures had not been previously employed to avoid lattice relaxation and interface smearing. It is important here that the thickness of the quantum well is much thinner than the critical thickness under thermal equilibrium conditions. This means that one can raise the growth temperature without worrying about lattice relaxation. At higher temperatures, atoms will move around on the surface and lattice imperfections such as missing atoms and disorder due to atoms displaced from the lattice sites may be effectively eliminated from the surface. It is worth mentioning that although Schäffler *et al* [7] reported that even low growth temperatures can provide QWS giving rise to edge emissions after improvement of the growth conditions, the luminescence efficiency increases with increasing growth temperature.

The heterointerface formed by high-temperature MBE was, however, found to be smeared due to surface

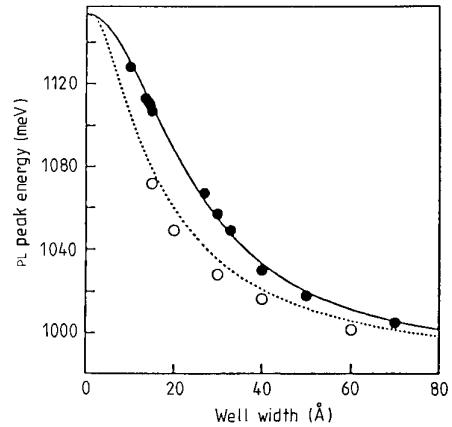


Figure 2. Well width dependence of PL peaks of $\text{Si}_{0.80}\text{Ge}_{0.20}/\text{Si}$ QWS grown by SAG (open circles) and conventional MBE at high temperatures (full circles). Full curve: calculation with segregation; dotted curve: calculation with square potential.

segregation. The full circles in figure 2 show the well width dependence of QW PL peaks [10]. It is seen that the peaks shift to higher energies from the positions calculated based on the square well; these positions are indicated by the dotted curve in the figure. This is because the quantum levels are lifted by the deformation of the well shape due to Ge surface segregation. This surface segregation was able to be described in terms of the two-state exchange model where the site exchange between Ge atoms in the subsurface state and Si atoms sitting in the surface state is taken into account [11–13]. The full curve in the figure is the well width dependence of QW edge emissions calculated by taking into account the distortion of the well shape based on the two-state exchange model. It is seen that the energy position is sufficiently represented by the calculation, indicating that the model adequately describes the surface segregation, though it is a phenomenological model and the real situation of atom exchange is not taken into account.

3. Segregant-assisted growth (SAG)

Copel and his co-workers developed the surfactant-mediated growth technique to suppress islandation of epitaxial films and demonstrated the formation of Ge layers with flat surfaces on Si substrates [14]. Their work gave rise to the idea of using strong segregants as suppressors of Ge segregation in overgrown Si layers; this process was named segregant-assisted growth (SAG) [11]. The principle of the SAG method is as follows: before the Si overlayer of the Si/SiGe/Si heterostructure is grown, the SiGe layer is covered with strong segregants like Sb so that Ge is no longer the topmost surface atom. When Si atoms are deposited on the Sb-covered Ge surface, the position of Sb becomes the subsurface state and Sb exchanges this position with impinging Si atoms sitting in the surface state. However, site exchange between Si and Ge atoms does not follow. This is because

Ge atoms occupy neither the surface state nor the subsurface state, that is, they are in the bulk state once Si is deposited and therefore they do not participate in surface segregation.

The open circles in figure 2 are energy positions of PL peaks of SAG samples as a function of well width. As can be seen in this figure, PL peaks of SAG samples are located at an energy lower than those of MBE samples and the experimental data are very close to the dotted curve calculated assuming an ideal square potential. Therefore, SAG can be said to effectively suppress surface segregation and provide abrupt interfaces.

In SAG growth, the number of segregant atoms required to suppress surface segregation is important; it was found that the effect becomes pronounced at around 0.5 ML and that 0.75 ML is the optimum in the case of Sb. It is also important to know which elements can act as segregation suppressors; so far As, Ga and Bi [15], apart from Sb, are known to effectively suppress Ge segregation.

Although segregant atoms mostly float up to the growing surface, some of the atoms were found to be incorporated into the overlayer. In the case of Sb atoms, moreover, it was found that they are selectively incorporated into Si layers rather than into Ge layers and that self-modulation doping is realized in SiGe/Si multiple QWs [16, 17]. The modulation doping is enhanced with increasing Ge content in SiGe alloy layers. This phenomenon was shown also to be well understood in terms of the above-mentioned two-state exchange model. This selective doping structure is advantageous both for the study of electronic states in quantum wells and for far-infrared detector applications. In fact, we succeeded in observing for the first time the optical absorption due to the intersubband transition in the conduction band of the extremely narrow QW fabricated on Ge substrates [18].

It is well known that type-2 QWs are formed when symmetrically strained SiGe/Si heterostructures are fabricated on SiGe buffer layers and that the Si layer becomes the well while the SiGe layer acts as the barrier. Since selectively doped Sb atoms act as donors, providing electrons in Si well layers, far-infrared light absorption due to the electronic transition from the ground state to the excited state in the quantum well, i.e. intersubband transition, becomes possible. Figure 3 shows the absorption spectra of several SiGe/Si QWs grown here and the spectral blue shift is clearly observed as the well width is decreased. The well width dependence of the absorption peak energy was found to approximately agree with the calculated energy difference between the ground and the first excited states when we assumed that the ideal square potential well was realized. It was also found that the observed peaks are systematically shifted about 10 meV above the calculated energies. This energy deviation of infrared light absorption may come from many-body effects of electrons and/or the deformation of the potential profile. It has been shown that the main contribution to the energy shift from the many-body effects is the electron-electron exchange interaction and that it lowers

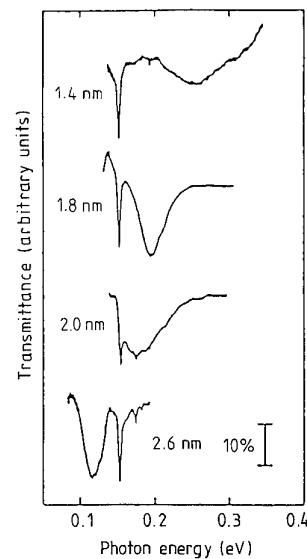


Figure 3. Far-infrared transmission spectra at 100 K of n-type $\text{Si}_{0.40}\text{Ge}_{0.60}/\text{Si}$ QWs (type-2 QWs) with various well widths grown on relaxed $\text{Si}_{0.70}\text{Ge}_{0.30}$ buffer layers.

the ground state, resulting in the peak shift to higher energies of about 16 meV for a well with $L = 1.4$ nm and electron concentration of $2 \times 10^{18} \text{ cm}^{-3}$. On the other hand, it was shown from the numerical calculation that the interfacial smearing due to Ge segregation lowers the intersubband transition energy and that the shift becomes a maximum when the Ge segregation length is comparable with the well width. The energy shift due to surface segregation becomes as high as 30 meV in the case of a QW of width 1.4 nm. Therefore, it can be concluded that the Ge surface segregation is well suppressed, to less than 0.4 nm, by SAG.

4. Gas-source molecular beam epitaxy (GSMBE)

Another important example of SAG, or pseudo-SAG, is gas-source MBE (GSMBE) where hydrogenated gases such as Si_2H_6 and GeH_4 are used as molecular beam sources. QWs grown by this method were found to give rise to intense band edge emissions [19, 20]. Figure 4 shows PL spectra with various well widths [19]. As seen in this figure, there are two main peaks, NP and TO, characteristic of SiGe layers, and they shift to higher energies as the SiGe well width is decreased. The well width dependence of the NP peak energies is plotted in figure 5. It should be noted that the well width dependence of the GSMBE sample coincides well with the values obtained by the square-well potential calculation. That is, GSMBE provides heterostructures without any interfacial smearing. This is probably due to the effect of a large amount of atomic hydrogen [21] which is constantly supplied to the growth front in this method. Recently, a preliminary experiment involving the introduction of atomic hydrogen into the solid-source MBE chamber has shown that surface segregation is suppressed even at temperatures above 500 °C

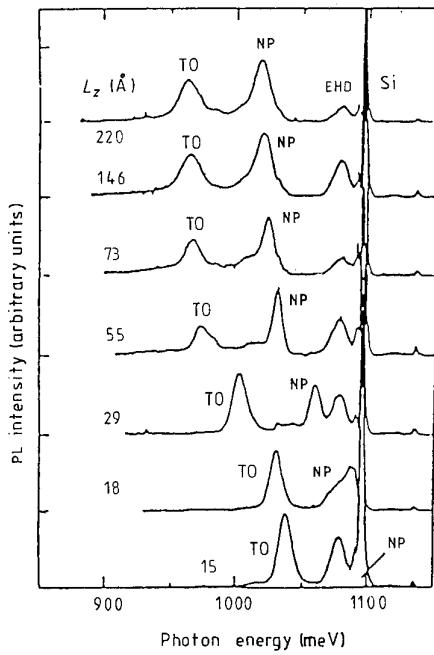


Figure 4. Well width (L_z) dependence of PL spectra at 18 K of $\text{Si}_{0.84}\text{Ge}_{0.16}/\text{Si}$ QWS grown by gas-source MBE.

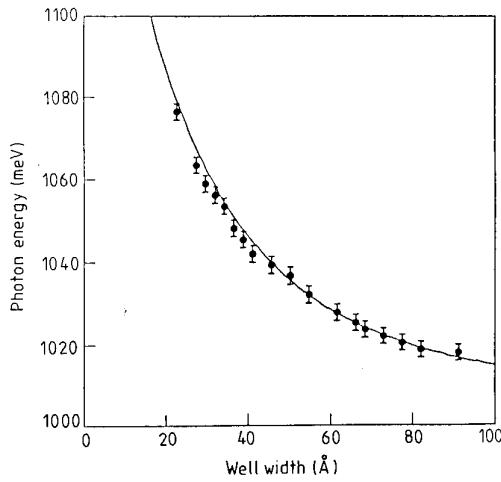


Figure 5. Well width dependence of NP line peak positions at 19 K in $\text{Si}_{0.82}\text{Ge}_{0.18}/\text{Si}$ QWS grown by gas-source MBE ($E_{\text{ex}} = 15 \text{ meV}$).

where hydrogen is considered to desorb from Si surfaces, supporting our speculation [22]. This probably suggests that the hydrogen has a relatively long residence time on Si surfaces, and it can then stay to act as a surfactant in the period of monolayer growth. It is also speculated that hydrogen may not only suppress the segregation but also terminate dangling bonds if they exist in epitaxial films.

It is surprising that SiGe/Si QWS consisting of indirect-bandgap materials provide highly efficient luminescence. This is considered to result from the fact that both electrons and holes are well confined in the same region, that is, the quantum well region, and radiative recombination is enhanced. This enhancement, however,

is realized, provided that the non-radiative centres are sufficiently eliminated from the quantum well region. The growth method developed here may realize such a situation. The temperature dependence of the integrated PL intensity of SiGe/Si QWS is shown in figure 6 [23]. It is seen that the intensity does not decrease with increasing temperature up to 100 K and that above 100 K it decreases with an activation energy of about 130 meV, which is approximately equal to the theoretically predicted band discontinuity in the valence band [24]. This fact supports the theoretical predictions [24] that the type-1 QW is formed and that the band offset almost goes to the valence band side and the conduction band discontinuity is negligibly small. Moreover, one can say that what controls the QW emissions is the confinement of the holes and not the electrons and that if the band discontinuity at either conduction band, not both, is sufficiently large, the quantum confinement of excitons is realized.

5. Luminescence from type-2 QWS and no-phonon lines

Another interesting subject is the formation of type-2 SiGe/Si heterostructures by gas-source MBE and their luminescence properties. To realize type-2 band alignment, Si layers should be under lateral tensile strain. The basic requirement for the growth of high-quality type-2 QWS and superlattices is formation of high-quality completely relaxed SiGe buffer layers. Figure 7 shows PL spectra of SiGe/Si(strained)/SiGe quantum wells and SiGe buffer layers with step-graded Ge concentration [25]. Two important features can be seen in this figure. First, the buffer layer (figure 7(f)) provides sharp and intense band edge emissions although it is relaxed and many misfit dislocations are accommodated at the interface between the buffer layer and Si substrate. This is a clear evidence for the high quality of the top region of the buffer where QWS are grown. Broad peaks are identified as dislocation- and defect-related emissions originating from the region close to the interface. The second feature is that additional peaks are seen in the case of QWS (a)–(e) grown on the buffer layer. This is quite interesting since the band alignment is spatially indirect and then electrons and holes are separately confined in Si and SiGe layers respectively. It was revealed that the energy position of the peaks is quite reasonable when we consider the spatially indirect transition between electrons confined in the Si well with square potential shape and holes in the SiGe barrier layer. It is also noteworthy that the quantum confinement effect is clearly seen, as indicated by Xs in figure 7.

Next we discuss the intensity ratio between the no-phonon (NP) peak and its phonon replica (TO) in QWS. As mentioned before, the NP emission becomes allowed in SiGe alloys since they have a random distribution of Si and Ge atoms. Therefore it is known that the NP peak becomes a maximum around 50% Ge content in bulk SiGe alloys [9] since the randomness is highest at this composition. In the case of type-1 QWS, the relative

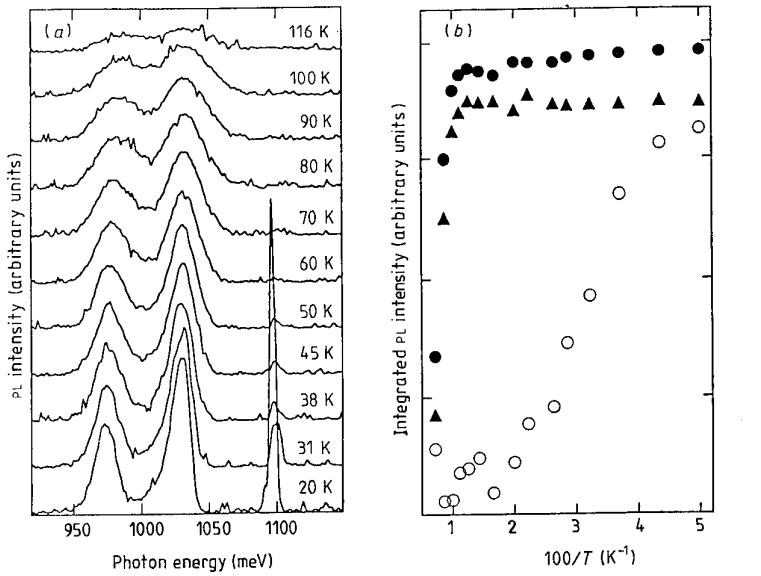


Figure 6. (a) Temperature dependence of PL profiles. (b) Integrated intensity of $\text{Si}_{0.84}\text{Ge}_{0.16}/\text{Si}$ QWs. (\circ , Si; \bullet , SiGe (NP); \blacktriangle , SiGe (TO)).

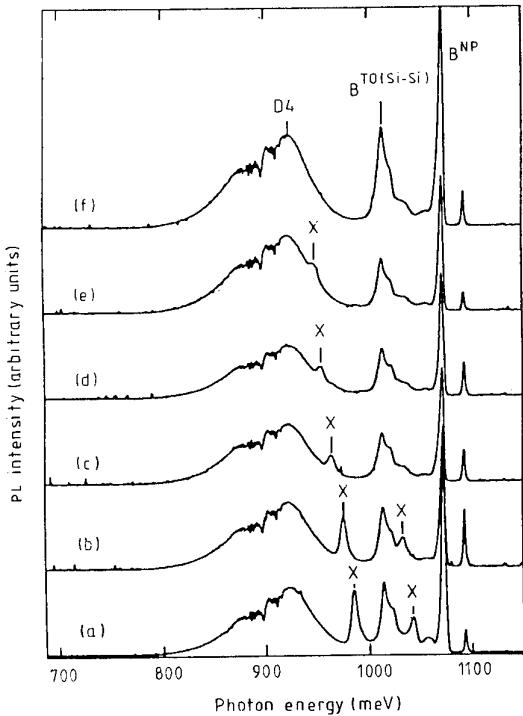


Figure 7. PL spectra at 4.2 K of (f) fully relaxed $\text{Si}_{0.82}\text{Ge}_{0.18}$ buffer layers and (a)–(e) type-2 QWs grown on the buffer with well widths of 10, 13, 16, 19 and 21 \AA respectively.

intensity of the NP peak is large in the wider wells and decreases with decreasing well width as seen in figure 5. This is because the wavefunctions of electrons and holes are well confined in the alloy layer and excitons fully experience the compositional randomness when the well width is large, resulting in a strong contribution of the NP transition to the spectrum. However, as the well width is decreased, the decay constant of the evanescent

wavefunction increases in the Si barrier layer, where an alloy contribution cannot be expected. This leads to the decrease of the NP line when the well width is narrow. On the other hand, the situation is different in type-2 QWs. That is, electrons are confined in the Si layer and only holes are located in the SiGe barrier layer, where alloy disordering plays a role. This means that the NP component in type-2 QWs is basically small compared with phonon-assisted modes. When the well width is decreased in type 2 QWs, the wavefunction tends to penetrate into the SiGe barrier layer where NP transition is allowed. The experimental result shows this situation well and the NP component increases with decreasing well width as seen in figure 7.

6. Formation of coupled wells and superlattices

The coupling of QWs, which is the simplest element of superlattices, is well understood based on the effective-mass approximation by precisely taking into account the band alignment. Figure 8 shows PL spectra of symmetrically coupled QWs with various barrier widths grown on Si substrates [26]. As can be seen in this figure, the peaks shift to higher energies with increasing barrier width. The peak energies of QWs whose barrier widths are zero and large correspond to a single QW (sqw) with double well width and two uncoupled QWs respectively, and their energies agree well with the calculated results as shown in figure 9. In this figure, the full curve is the calculated barrier width dependence. The good agreement over the whole range indicates that the theoretical prediction that the band alignment is type-1 when the Ge content is lower than 0.6 is correct and that ideally square-shaped wells are formed by this method.

Figure 10 shows the energy shifts of PL peaks when

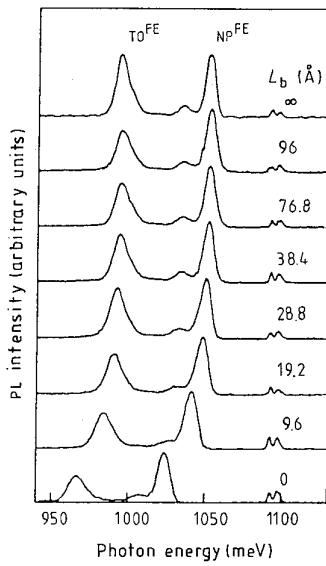


Figure 8. PL spectra at 19 K of symmetrically coupled $\text{Si}_{0.82}\text{Ge}_{0.18}/\text{Si}$ QWS with various barrier widths (well width, L_z , is 34 Å).

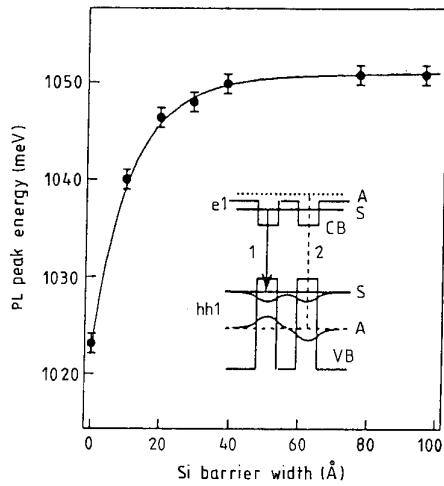


Figure 9. NP peak energies at 19 K of coupled $\text{Si}_{0.82}\text{Ge}_{0.18}/\text{Si}$ QWS with well width L_z of 34 Å as a function of barrier layer width.

the number of coupled wells is increased, that is, the evolution of superlattices [26]. Since the increase in the number of coupled wells lowers the ground state energy and finally forms a miniband, the PL peak energy, corresponding to the band edge, decreases with increasing well number. This situation is clearly seen in this figure and an effective-mass calculation was found to give reasonable values for the energy shift.

7. Electroluminescence of SiGe/Si QWS

By using the SAG method, electroluminescent (EL) diodes were successfully fabricated [27, 28]. Figure 11 shows the EL spectrum along with the PL spectrum of the diode

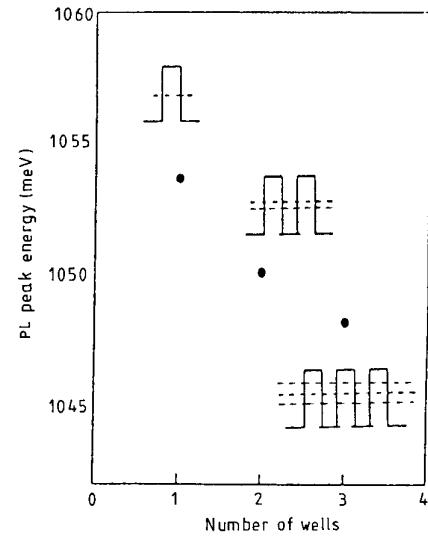


Figure 10. PL peak energies as a function of number of coupled $\text{Si}_{0.82}\text{Ge}_{0.18}/\text{Si}$ QWS.

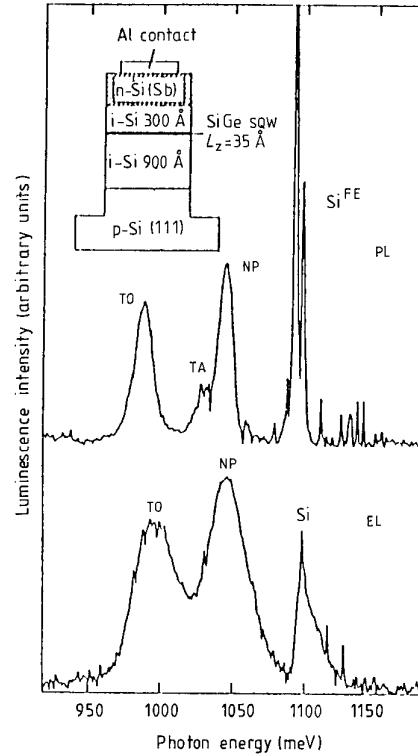


Figure 11. PL (at 18 K) and EL (at 20 K) spectra of $\text{Si}_{0.80}\text{Ge}_{0.20}/\text{Si}$ QWS grown on an Si (111) substrate.

fabricated. This diode has a single QW and was grown on p-type Si (111) substrates at 700 °C. The reason why (111) orientation is employed is that the band discontinuity at the valence band is larger on Si (111) surfaces than those of other orientations; that is, the discontinuity at the conduction band is absent in this direction. It is seen that almost the same spectrum is observed in both EL and PL and that EL is sufficiently intense. It is noted, however, that EL peaks are broader than PL peaks and

that the relative intensity of QW emissions to Si bulk emissions is different between EL and PL spectra. This is due to the fact that the carrier density and distribution are different between them. That is, carriers are highly injected mostly in the QW region in the case of EL, causing broadening of the emission peaks. On the other hand, photoexcitation generates carriers widely in the whole sample including the Si substrate, resulting in the increase of relative intensity of Si-related emissions. By taking this situation into account, it was demonstrated that EL diodes which operate even at room temperature can be fabricated [27].

8. Formation of quantum wires and their optical properties

The wire geometry was realized by growing SiGe/Si quantum wells on Si (100) substrates V-groove-patterned with (111) facets. It is known that selective growth takes place between Si and SiO_2 substrates when GSMBE is employed and quantum structures were formed only on the Si surfaces [29, 30]. In the bottom of the V-groove, crescent-shaped SiGe features, corresponding to quantum wires, were found to have grown [31, 32]. The lateral and vertical wire sizes were approximately 300 and 80 Å respectively, as observed by TEM. Figure 12 shows PL spectra of quantum structures grown (a) on a planar Si (100) substrate and (b) on a V-groove-patterned Si substrate. For the control QW the well width and Ge content are nominally 34 Å and 0.18 respectively. It is significant that a large blue-shift is seen in the spectrum of the latter sample. Since it is confirmed that very thin SiGe layers which do not give rise to luminescence were grown on (111) V-grooved surfaces under the growth

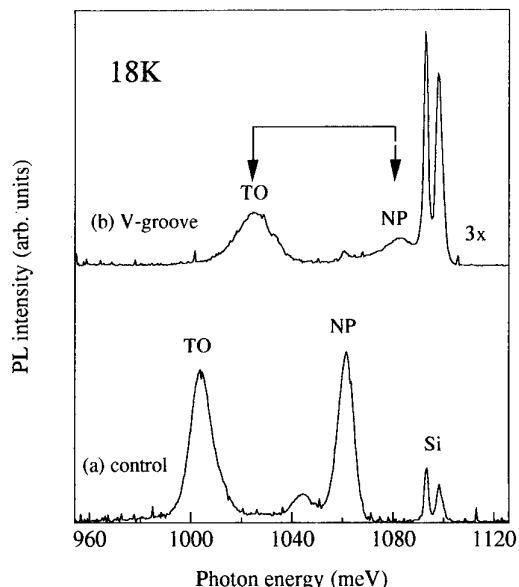


Figure 12. PL spectra of QWRs grown on V-groove Si (100) substrates by gas-source MBE along with those of $\text{Si}_{0.82}\text{Ge}_{0.18}/\text{Si}$ QWS with the same well width.

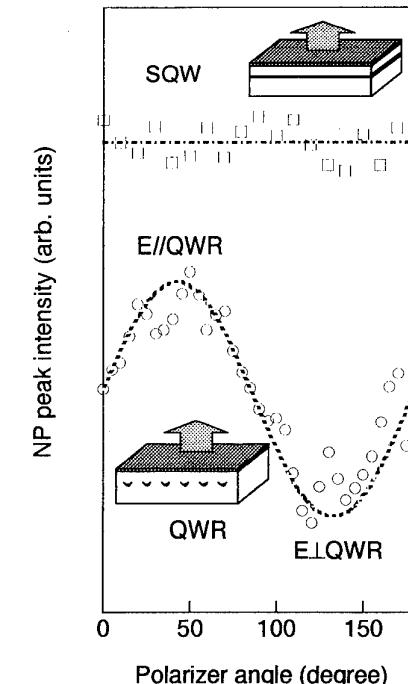


Figure 13. Polarization of luminescence from QWS and QWRs of the samples shown in figure 12.

conditions employed here, the luminescence can be assigned to the contribution from the wire structure. To confirm this, polarization measurements were carried out and the result is shown in figure 13. It is remarkable that the polarization in these structures is different. That is, no polarization is seen in the cross-sectional emission while strong polarization is observed in the emission from the surface in the case of QWR structures. This is in strong contrast to the polarization characteristics of QWS as seen in figure 13. This result indicates that the luminescence definitely comes from the wire structures in the V-grooved sample and luminescent QWR-like structures are realized. However, the energy shift in QWRs is too big and does not agree with a simple estimation based on the size. This situation is different from the case of AlGaAs/GaAs QWRs on a V-grooved GaAs substrate where the energy shift is reasonable. In the case of SiGe QWRs, strain relief which widens the bandgap and/or the spatial variation of Ge composition is likely to occur. This would cause the unexpectedly large energy shift.

9. Summary

High-quality SiGe/Si QW structures were successfully formed by a segregant-assisted growth technique. This enables us to perform investigations of scientifically important subjects such as band alignment, exciton trapping, transport of photogenerated carriers in indirect materials, electron-electron interactions in QWS and so on. New quantum structures such as quantum wires are also realized by this method and quantum phenomena

relating to the lower dimensionality are now being exploited. Moreover, it should be pointed out that optical devices such as infrared light detectors and light emitting diodes can be fabricated, suggesting the high potential of SiGe/Si heterostructures for device application.

Acknowledgments

This work was done in collaboration with N Usami, K Fujita, T Mine and D K Nayak, and supported in part by a Grant-in-Aid from the Japanese Ministry of Education, Science and Culture.

References

- [1] Harris J J, Ashenford D E, Foxon C T, Dobson P J and Joyce B A 1984 *Appl. Phys. A* **33** 87
- [2] Metzger R A and Allen F G 1984 *J. Appl. Phys.* **55** 931
- [3] Sturm J C, Manoharan H, Lenchyshyn L C, Thewalt M L, Rowell N L, Noël J-P and Houghton D C 1991 *Phys. Rev. Lett.* **66** 1362
- [4] Noël J-P, Rowell N L, Houghton D C and Perovic D D 1990 *Appl. Phys. Lett.* **57** 1037
- [5] Usami N, Fukatsu S and Shiraki Y 1992 *Appl. Phys. Lett.* **61** 1706
- [6] Brunner J, Menczigar U, Gail M, Friess E and Abstreiter G 1992 *Thin Solid Films* **222** 27
- [7] Schäffler F, Wachter M, Herzog H-J, Thonke K and Sauer R 1993 *J. Crystal Growth* **127** 411
- [8] Thompson P E, Godbey D, Hobart K, Glaser E, Kennedy E, Twigg M and Simons D 1993 *Int. Conf. on Solid State Devices and Materials (Chiba, 1993)*, extended abstracts p 234
- [9] Weber J and Alonso M I 1989 *Phys. Rev. B* **40** 5683
- [10] Usami N, Fukatsu S and Shiraki Y 1993 *Appl. Phys. Lett.* **63** 388
- [11] Fujita K, Fukatsu S, Yaguchi H, Igarashi T, Shiraki Y and Ito R 1990 *Japan. J. Appl. Phys.* **29** L1981
- [12] Fujita K, Fukatsu S, Yaguchi H, Shiraki Y and Ito R 1991 *Appl. Phys. Lett.* **59** 2240
- [13] Fukatsu S, Fujita K, Yaguchi H, Shiraki Y and Ito R 1991 *Appl. Phys. Lett.* **59** 2103
- [14] Copel M, Reuter C, Kaxiras E and Tromp R M 1989 *Phys. Rev. Lett.* **63** 632
- [15] Sakamoto K, Kyoya K, Miki K, Matsuhara H and Sakamoto T 1993 *Japan. J. Appl. Phys.* **32** L204
- [16] Fujita K, Fukatsu S, Usami N, Yaguchi H, Shiraki Y and Ito R 1993 *Proc. 5th Int. Conf. on Shallow Impurities in Semiconductors (Kobe, 1993) Mater. Sci. Forum* **117 & 118** p 153
- [17] Fujita K, Fukatsu S, Yaguchi H, Shiraki Y and Ito R 1993 *J. Crystal Growth* **127** 416
- [18] Fujita K, Fukatsu S, Shiraki Y, Yaguchi H and Ito R 1992 *Appl. Phys. Lett.* **61** 210
- [19] Fukatsu S, Yoshida H, Fujiwara A, Takahashi T, Shiraki Y and Ito R 1992 *Appl. Phys. Lett.* **61** 804
- [20] Terashima K, Tajima M, Ikarashi N, Niino T, Hiroi M and Tatsumi T 1992 *J. Electron. Mater.* **21** 1081
- [21] Copel M and Tromp R M 1991 *Appl. Phys. Lett.* **58** 2648
- [22] Ota G, Usami N, Fukatsu S and Shiraki Y *Appl. Phys. Lett.* to be published
- [23] Fukatsu S, Yoshida H, Usami N, Fujiwara A, Takahashi T, Shiraki Y and Ito R 1992 *Japan. J. Appl. Phys.* **31** L1319
- [24] Van de Walle C G and Martin R M 1986 *Phys. Rev. B* **34** 5621
- [25] Nayak D K, Usami N, Fukatsu S and Shiraki Y 1993 *Appl. Phys. Lett.* **63** 3509
- [26] Fukatsu S and Shiraki Y 1993 *Int. Conf. on Solid State Devices and Materials (Chiba, 1993)*, extended abstracts p 895
- [27] Fukatsu S, Usami N and Shiraki Y 1993 *Japan. J. Appl. Phys.* **32** 1502
- [28] Fukatsu S, Usami N, Chinzei T, Shiraki Y, Nishida A and Nakagawa K 1992 *Japan. J. Appl. Phys.* **31** L1015
- [29] Hirayama H, Tatsumi T and Aizaki N 1988 *Appl. Phys. Lett.* **52** 2242
- [30] Hirayama H, Hiroi M, Koyama K and Tatsumi T 1990 *Appl. Phys. Lett.* **56** 1107
- [31] Usami N, Mine T, Fukatsu S and Shiraki Y 1993 *Appl. Phys. Lett.* **63** 2789
- [32] Usami N, Mine T, Fukatsu S and Shiraki Y 1994 *Appl. Phys. Lett.* **64** 1126

Optical processes in microcavities

R E Slusher

AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

Abstract. Optical microresonators can now be fabricated with dimensions of the order of half a wavelength on a side. For semiconductor microresonators this results in only a few optical modes interacting with the emitting material in the cavity. In this limit the threshold lasing characteristics are dramatically modified. Semiconductor microdisc experiments will be described where as much as 20% of the spontaneous emission is captured into the single lasing mode. This strong coupling of the radiation field with the resonant mode can cause anomalous laser linewidths and fluctuation phenomena. These experiments provide a good test of detailed models of dense, non-equilibrium electron-hole gases. Extensions of optical microcavities to a broad range of material systems will also be discussed.

1. Introduction

Semiconductor fabrication techniques have advanced to the stage where it is possible to make optical resonators with dimensions of the order of an optical wavelength [1] in the material. In this microcavity limit there is only one low-loss mode that interacts with the optically active material in the cavity. In addition to the very-low-threshold lasers that can be obtained with these microcavities there are many interesting phenomena associated with microcavities including control of spontaneous emission rates, anomalous laser linewidth behaviour, and non-equilibrium effects in the active region. We have explored microcavity resonators based on the low-loss modes propagating near the edge of any curved dielectric surface. These ‘whispering-gallery’ resonances have been used for lasers since the early work of Garrett *et al* [2] who used small spheres of ruby. Richard Chang and his collaborators pioneered microcavity effects in nearly spherical liquid droplets [3]. We will describe semiconductor microcavities [4–7] formed into disc and cylindrical shapes by photolithographic patterning and selective etching techniques. These techniques may be applicable to a wide range of material systems since low optical cavity loss is obtained from a simple curved surface instead of multiple layer dielectric mirrors. For example, dye-doped polymer sphere [8] and II–VI semiconductor discs [9] have recently been demonstrated as lasers.

From a basic physics point of view the microdisc lasers allow us to study the behaviour of lasers in interesting limits. For example, in the limit of resonator volumes near a cubic wavelength in the material a large fraction β of the spontaneous emission from the optically active region is emitted into the lasing modes. We have measured β values as high as 0.2 for semiconductor discs that are 2 μm in diameter and 0.15 μm thick. This strong coupling of the radiation field to the gain medium can cause strong non-equilibrium effects and dramatic modifications of the laser threshold characteristics. In

effect we have a unique microlaboratory for the study of lasing phenomena and the hot, dense electron-hole gas in the gain region.

2. Microdisc resonators

A cross section of a semiconductor microdisc is shown in figure 1 for the InGaAsP/InP material system. The dominant mode of this disc is a transverse electric (TE) mode that propagates within a few tenths of a micrometre of the disc edge. We will neglect the effects of the support post, although it can be important for optical losses if it is too close to the disc edge. The disc thickness is less than half a wavelength in the material in order to eliminate higher-order TE modes and all TM modes. A small fraction of the mode extends outside the disc boundary so that the effective index of refraction, n_{eff} , is less than the bulk index. For example, for the disc shown in figure 1 with a thickness of 0.15 μm , $n_{\text{eff}} = 2.35$. The scalar optical field ψ in the disc can be approximated by solutions of the two-dimensional Helmholtz equation

$$\psi = J_M(x) e^{iM\theta} \quad (1)$$

where $x = n_{\text{eff}}(\omega)\omega r/c$, ω is the angular optical frequency, J_M is a Bessel function of order M and r and θ are the polar coordinates of the disc. The lowest-order radial mode has the lowest loss and has a single field maximum just inside the disc radius. The radial modes are assigned the mode number $N = 1, 2, \dots$ and the mode frequencies are $\omega_{M,N}$ with $\omega_{M,1} = \omega_M$. It is a good approximation to set $\psi = 0$ as a boundary condition at the disc radius R . Then the solution to equation (1) is found from $n_{\text{eff}}(\omega)R\omega/c = x_M^{(1)}$ is the smallest positive root of $J_M(x_M^{(1)}) = 0$.

Optical losses in the disc resonator are characterized by the quality factor Q , the number of radian cycles required for field energy to decay by a factor of $1/e$. There are three major contributions to Q

$$Q^{-1} = Q_m^{-1} + Q_a^{-1} + Q_i^{-1} \quad (2)$$

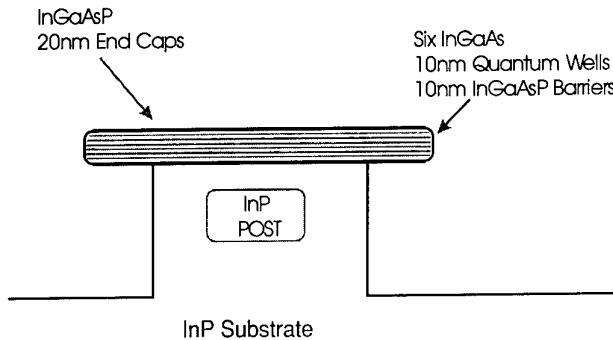


Figure 1. An InGaAs/InGaAsP microdisc is shown schematically in cross section. It has a designed thickness of $0.15 \mu\text{m}$ and contains six 10 nm thick $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ quantum wells separated by five 10 nm thick InGaAsP barriers. The disc is enclosed by two 20 nm thick InGaAsP protective end caps.

where

$$Q_m = b e^{2MJ} \quad (3)$$

is the radiation loss from the ideal disc and

$$J = \operatorname{arctanh} \left[\left(1 - \frac{1}{n_{\text{eff}}^2} \right)^{1/2} \right] - \left(1 - \frac{1}{n_{\text{eff}}^2} \right)^{1/2}. \quad (4)$$

The prefactor b is approximately 1/7. This radiative loss should also be corrected [4] for the relatively large dispersion of n_{eff} . Optical absorption losses are included in the Q_a term and Q_i includes losses due to scattering from roughness at the disc edges and the support post. Another important contribution to Q_i may be loss from lower-order modes coupled into the dominant lasing mode by dielectric perturbations with periods of the order of the wavelength in the disc. In practice the dominant loss for pump powers near and above the lasing threshold is attributed to Q_i since Q_m can be very large for disc radii larger than $1.5 \mu\text{m}$ for the high-index semiconductors. Measured values of Q range from 200 to over 1000.

The radiation pattern from the ideal disc is directed out radially in a relatively narrow range of angles above and below the plane of the disc. Experimentally there is appreciable radiation in the vertical direction not predicted by the ideal model. It is likely that all of the radiation coupled out of the lasing discs is due to dielectric roughness at the edges or the post geometry. We have not made a thorough study of the radiation patterns to date.

3. Experimental results

Experimentally we have concentrated on two material systems, the InP/InGaAsP system as shown by the example in figure 1 and the GaAs/AlGaAs system shown in the schematic diagram in figure 2. The disc radii are in the range between 0.8 and $5 \mu\text{m}$ and the disc thicknesses range between 0.1 and $0.2 \mu\text{m}$. After growing the layered semiconductor wafer by MOCVD techniques a photoresist pattern defines the discs which are then etched

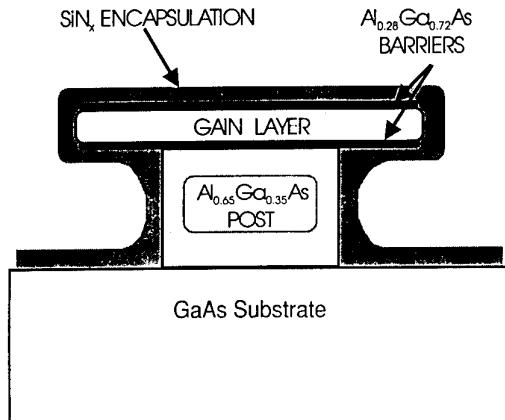


Figure 2. A schematic diagram of the cross section of a GaAs/AlGaAs microdisc laser. The gain layer is either bulk GaAs or 5 nm GaAs quantum wells. The post height is $0.5 \mu\text{m}$. Barriers of $\text{Al}_{0.28}\text{Ga}_{0.72}\text{As}$ on either side of the gain layer confine the optically excited carriers to the gain region. The silicon nitride deposition is approximately 40 nm thick and is shown as a dark grey encapsulation layer.

into cylinders by electron cyclotron etching. A selective chemical etch, HCl for the InP and HF for the AlGaAs, removes a small ring of material under the disc to leave the edge of the disc exposed to air.

Non-radiative recombination at the edge of the GaAs active regions reduces the optical gain to very low levels unless the surface is passivated with a sulphur treatment. The resulting thin passivation layer must be covered with a protective layer of silicon nitride to prevent photodesorption of the sulphur as shown in figure 2. The discs can be optically pumped by a HeNe laser or a pn junction can be formed above and below the disc for electrical pumping [7]. A series of luminescent spectra from a $1 \mu\text{m}$ radius GaAs microdisc are shown in figure 3 for pump powers near the lasing threshold. Note that well below threshold there is a cavity-enhanced spectral feature approximately 0.8 nm wide that is a measure of the cavity $Q = 1000$. The threshold pump power is $40 \mu\text{W}$. The GaAs gain region for these experiments is bulk material with a thickness of $0.12 \mu\text{m}$. GaAs quantum wells in the gain region result in similar thresholds.

Spectra for the InGaAs quantum well discs with radii of $1.1 \mu\text{m}$ and $2.5 \mu\text{m}$ obtained by optical pumping at liquid-nitrogen temperatures are shown in figures 4 and 5. Except for a few discs, there is only one lasing mode at all pump powers. In a few discs at the highest pump powers two modes lased allowing the mode spacings to be determined. The measured mode spacings are consistent with M values of 7 and 19 for the $1.1 \mu\text{m}$ and $2.5 \mu\text{m}$ InGaAs discs respectively.

The power in the lasing mode is shown as a function of optical pump power in figures 6 and 7 for the InGaAs microdisks. The threshold for lasing is $25 \mu\text{W}$ and $45 \mu\text{W}$ for the $1.1 \mu\text{m}$ and the $2.5 \mu\text{m}$ radii discs respectively. Note that the transition into the linear light output lasing region is more gradual for the smaller disc. This gradual transition to the lasing regime is predicted by the model calculation shown as the full curve in figure 7. This model

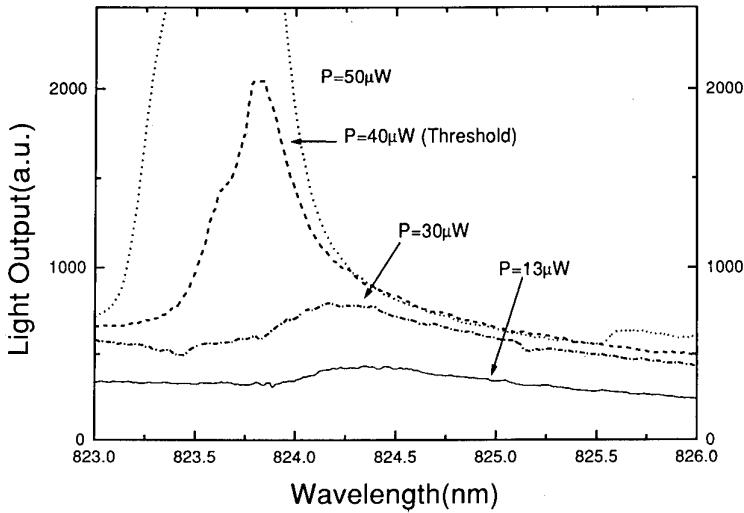


Figure 3. Photoluminescence spectra from a $1 \mu\text{m}$ radius GaAs/AlGaAs microdisc are shown for a series of pump powers near the lasing threshold. The discs are cooled to temperatures near liquid nitrogen and optically pumped with a HeNe laser. The full curve is well below threshold at a pump power of $13 \mu\text{W}$, the chain, broken and dotted curves are at pump powers of 30 , 40 and $50 \mu\text{W}$ respectively. A broad luminescent background from the GaAs band edge extends from approximately 820 nm to 830 nm . A Q value of 1000 is calculated from the full-width half-maximum of the enhanced luminescence due to the microdisc cavity seen in the portion of the full curve peaked at 824.3 nm .

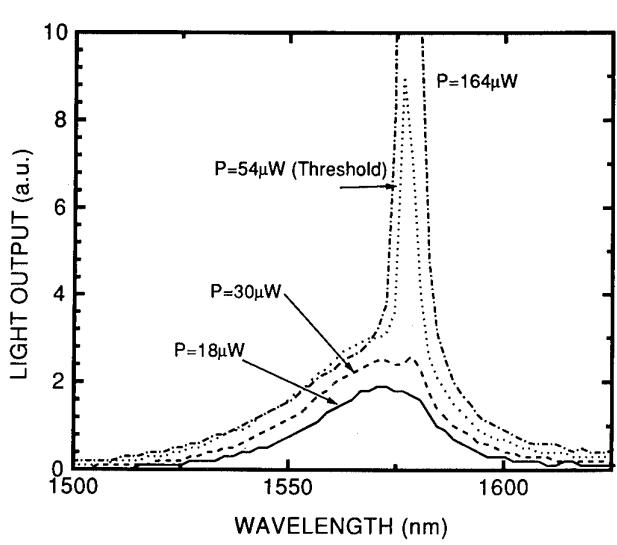


Figure 4. Photoluminescence spectra from a $5 \mu\text{m}$ diameter InGaAs/InGaAsP microdisc are shown for a series of pump powers near threshold. The pump powers are $18 \mu\text{W}$ (full curve), $30 \mu\text{W}$ (broken curve), $54 \mu\text{W}$ (dotted curve) and $164 \mu\text{W}$ (chain curve).

calculation uses rate equations to describe the carrier and photon number relaxation processes. It assumes quasi equilibrium Fermi distributions for both the electrons and holes. Both stimulated and spontaneous emission rates are adjusted by a fitting parameter β from which we estimate [5] that as much as 23% (15% at threshold) of the total spontaneous emission for the optically active region is captured into the lasing mode for this small

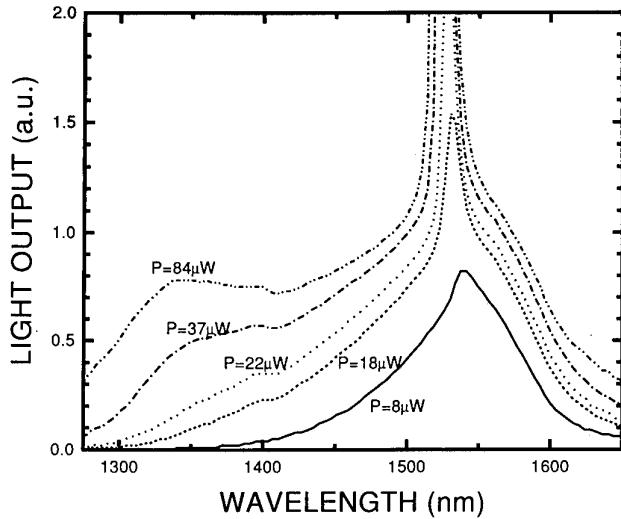


Figure 5. Photoluminescent spectra are shown for a $2.2 \mu\text{m}$ diameter InGaAs/InGaAsP microdisc for a series of pump powers in the range near threshold. The pump powers are $8 \mu\text{W}$ (full curve), $18 \mu\text{W}$ (broken curve), $22 \mu\text{W}$ (dotted curve), $37 \mu\text{W}$ (chain curve) and $84 \mu\text{W}$ (dash-double-dot curve).

disc. We are approaching the regime where it is more difficult to distinguish between spontaneous and stimulated emission.

The measured threshold pump powers in all of these experiments are in qualitative agreement with model calculations. The low threshold values result from the small volumes of gain material in the disc. Even at room

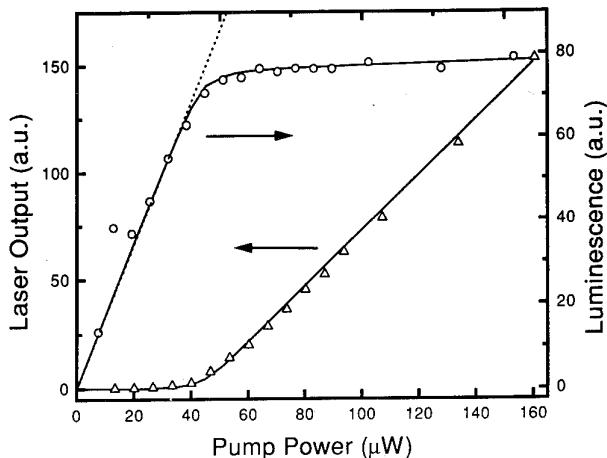


Figure 6. Light output at the laser line (triangles) and the integrated photoluminescence output (open circles) are shown as a function of pump power. The full curves are model calculations using $\beta = 0.23$ and $Q = 500$. The dotted line is the integrated photoluminescence from a non-lasing disc.

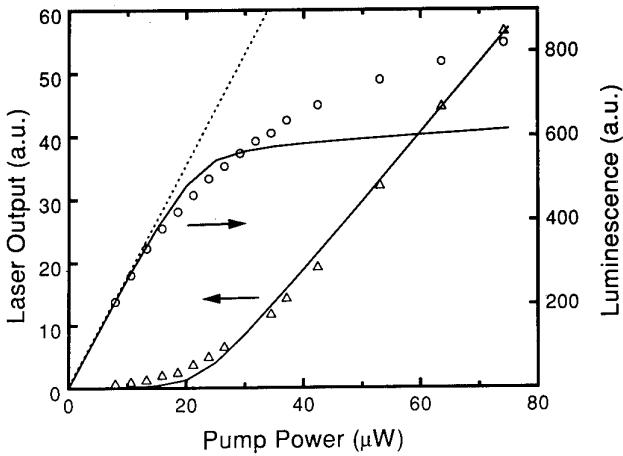


Figure 7. Light output at the laser line (triangles) and the integrated photoluminescence output (open circles) are shown as a function of pump power. The full curves are a model calculation using $\beta = 0.04$ and $Q = 800$. The dotted line is the integrated photoluminescence from a non-lasing disc. The pump powers used here and in figure 6 are the absorbed pump powers integrated over the entire area of the disc.

temperature threshold pump currents for electrically pumped structures have been measured [6] below 1 mA.

4. Non-equilibrium carrier dynamics in semiconductor microcavities

The simple mode structure and strong coupling between the radiation from the active region and the cavity mode make the semiconductor microcavity an interesting microlaboratory for the study of the hot, dense electron gas that comprises the gain region of most semiconductor lasers. As the β value approaches unity the stimulated emission rates become comparable to the energy equi-

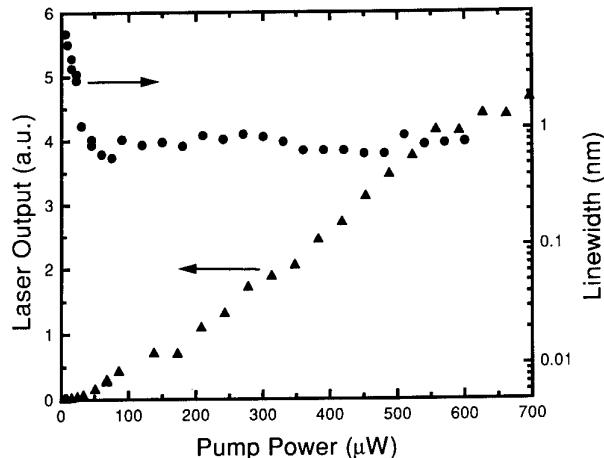


Figure 8. The output light at the laser line (full triangles) and the laser linewidth (full circles) are shown as a function of pump power for a $2.2 \mu\text{m}$ diameter InGaAs/InGaAsP microdisc cooled to temperatures near liquid nitrogen and optically pumped with an HeNe laser. The laser threshold is near $35 \mu\text{W}$.

libration rates in the lasing regime. This leads to non-equilibrium energy distributions, especially for the electrons. A comparison of figures 4 and 5 clearly shows these non-equilibrium effects. For the larger microdisc the luminescence spectrum and intensity (figure 4) remain constant at pump powers above threshold. This is also seen in the integrated luminescent output shown in figure 6. This is the normal behaviour of a lasing gain material. The carrier density and energy distribution are pinned at threshold since each additional electron–hole pair created at pump powers above threshold is almost immediately converted to a stimulated photon. In contrast to this typical behaviour, the luminescent spectral intensity continues to increase above threshold for the small-volume, high- β microdisc as seen in figures 5 and 7. Also note that the spectrum increases disproportionately at short wavelengths. The strong stimulated emission at large β is draining the electrons out of the conduction band faster than the energy relaxation processes can cool them from the higher energies where they are excited by the HeNe pump. The energy relaxation times are expected to be in the 1 to 10 ps range and are dominated by optic and acoustic phonon emission processes. Momentum relaxation times are much faster in the range near 50 fs. If the luminescence spectrum in figure 5 is compared with that of a non-lasing disc with similar dimensions, the lasing disc appears to have a spectral hole burned into it around the lasing energy with a width that corresponds to a relaxation time of 40 fs, in good agreement with the momentum relaxation time estimates.

We have studied the lasing linewidths of the microdisc lasers as a function of β and find that they have large linewidths, nearly independent of pump power above threshold as shown in figure 8. This is not the typical behaviour for semiconductor laser linewidths. The laser linewidth is expected to decrease inversely with the pump

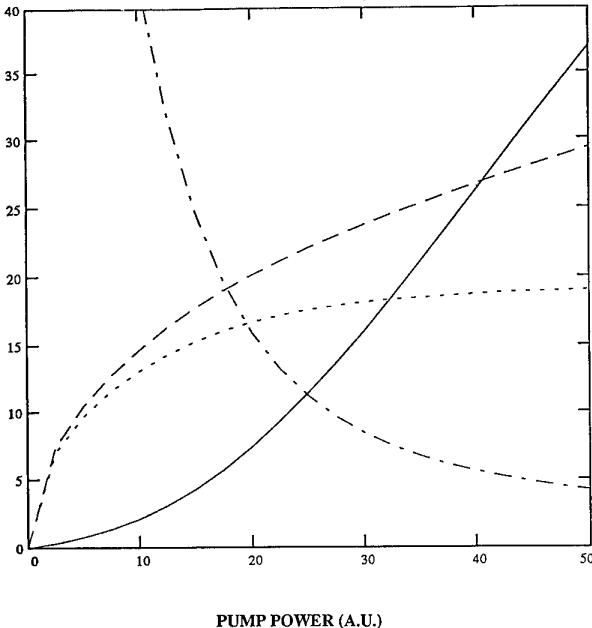


Figure 9. Three rate equations for hot electrons, electrons with energies near the lasing transition and the number of photons in the cavity mode yield the solutions shown here for the number of photons in the cavity mode (full curve), density of hot electrons (broken curve), density of electrons at the lasing energy (dotted curve) and the laser linewidth (chain curve). The left y axis is in units of photon number in the cavity mode and the other curves are in arbitrary units.

power to values in the 10 to 100 MHz range. Instead we see linewidths near 100 GHz that are independent of pump power for almost two decades above threshold. The inverse pump power dependence is expected since the linewidths are dominated by spontaneous emission noise which remains constant above threshold because of the carrier pinning discussed earlier. This constant noise along with the increasing light amplitude of the laser radiation leads to the normal ‘Schawlow–Townes’ linewidth behaviour. The high- β , constant-linewidth behaviour can be understood from the increasing spontaneous emission noise associated with the non-equilibrium carriers. The non-equilibrium increase in carrier density above threshold leads to a nearly linear increase in spontaneous emission noise resulting in the constant linewidths that we observe.

A simple model of this process can be set up to demonstrate these non-equilibrium effects. We use three rate equations to describe the ‘hot’ carriers, the carriers at the lasing energy and the photons in the cavity. The hot carriers relax to the carriers at the lasing energy with a characteristic energy relaxation time and also decay by spontaneous emission. The carrier density at the lasing energy increases only as a result of energy relaxation from the hot carriers. The results of this simple model are easily calculated numerically, and the pump power dependence of the carrier populations, the photon number and the linewidth are shown in figure 9. The energy relaxation time in the model was approximately 50 ps and the β

value was 0.2. In this case the energy relaxation and the stimulated emission rates are nearly equal when there are 50 photons in the cavity mode. The hot carrier densities as well as the carriers at the lasing energy continue to increase above threshold. This leads to linewidths (proportional to the carrier density at the lasing energy) that decrease at a pace less than that predicted by the Schawlow–Townes model. This simple rate equation model only serves to illustrate the physics involved. It requires an energy relaxation time that is appreciably longer than one might expect. Values near 10 ps are estimated for energy relaxation times at room temperature. These times are long compared with the optic phonon emission times near 1 ps. There are effects of phonon bottlenecks and screening of the electron–phonon interaction at high carrier densities that may appreciably lengthen the energy relaxation times. A much more detailed analysis of the non-equilibrium effects and the anomalous linewidths is in progress using non-equilibrium semiconductor Bloch equations [10].

5. Summary

Semiconductor microdiscs have been demonstrated that have Q values near 1000 for diameters as small as 2 μm . This corresponds to volumes of only a few cubic wavelengths and only a single high- Q mode in the gain spectrum. In these microcavities as much as 20% of the spontaneous emission is captured into the single lasing mode. The strong coupling of the emitted radiation to the microcavity resonant mode leads to non-equilibrium carrier dynamics and anomalous laser linewidths in the lasing regime. The microcavities serve as a micro-laboratory for the study of carrier dynamics which may help us to understand laser modelling over a wide range of temperatures. There are a number of other interesting phenomena in this microcavity regime. Experiments are in progress to measure the photon correlations near threshold in the large- β limit. At small β values there are large-amplitude fluctuations in a narrow range of pump powers near threshold, a characteristic of second-order phase transitions. At large β these fluctuations should decrease and spread over a wider range of pump powers.

The simplicity of the microdisc resonator may be useful for fabricating lasers in a wide range of materials where other resonator designs are difficult. For example, the vertical cavity microlasers using multilayer dielectric mirrors are very difficult to fabricate in the InP/InGaAsP system because the index of refraction differences available are too small. We are investigating wide-bandgap systems including organic materials and wide-bandgap semiconductors that emit in the visible range. It would be excellent if we could find more systems like liquid droplets that form extremely uniform surfaces. Recently glass droplets have been formed with Q values as high as 10^9 .

At present we have not demonstrated applications of the microdisc lasers. They have been demonstrated to operate at room temperature in low-repetition-rate

pulsed mode. However, the thermal contact to the substrate for the discs is roughly a factor of two worse than for the more common buried gain region designs used for vertical cavity and stripe semiconductor lasers. There is hope that minimization of the cavity losses and optimization of the cavity thermal contact will result in continuous room-temperature operation for microdisc geometries. Another major problem for the disc cavities is the mode coupling efficiencies to optical fibres or other photonic circuit elements. Very efficient coupling in the range from 50 to over 90% is required for many applications. For example, one important advantage of vertical cavity lasers is their excellent mode quality for coupling into optical fibres. In principle one can envision elegant photonic circuits based on cylinder or disc sources, amplifiers, nonlinear elements and waveguide couplers. The high dielectric contrast between the surrounding air or low-dielectric-constant encapsulating material and the high-dielectric-constant semiconductor discs and waveguides allows very dense and highly functional circuits due to the small bending losses, even for curvature radii of only a few wavelengths. It is not clear at this point whether photolithographic or self-assembly techniques will allow the precision control of

the dimensions required for this type of microphotonic circuitry.

References

- [1] Yamamoto Y and Slusher R E 1993 *Phys. Today* **46** 66
- [2] Garrett C G B, Kaiser W and Bond W L 1961 *Phys. Rev.* **124** 1807
- [3] Chen G, Chang R K, Hill S C and Barber P W 1991 *Opt. Lett.* **16** 1269
- [4] McCall S L, Levi A F J, Slusher R E, Pearton S J and Logan R A 1992 *Appl. Phys. Lett.* **60** 289
- [5] Slusher R E, Levi A F J, Mohideen U, McCall S L, Pearton S J and Logan R A 1993 *Appl. Phys. Lett.* **63** 1310
- [6] Levi A F J, Slusher R E, McCall S L, Tanbun-Ek T, Coblenz D L and Pearton S J 1992 *Electron. Lett.* **28** 1010
- [7] Hobson W S, Mohideen U, Pearton S J, Slusher R E and Ren F 1993 *Electron. Lett.* **29** 2199
- [8] Kuwata-Gonokami M, Takeda K, Yasuda H and Ema K 1992 *Japan. J. Appl. Phys. Lett.* **31** 99
- [9] Hovinen M, Ding J, Nurmiikko A V, Grillo D C, Han J, He L and Gunshor R L 1993 *Appl. Phys. Lett.* **63**
- [10] Mohideen U, Slusher R E, Jahnke F and Koch S W 1994 *Phys. Rev. Lett.* submitted

SECTION 4. WIRES: SPECTROSCOPY AND LASING ACTION

1933 Lasing in lower-dimensional structures formed by cleaved edge overgrowth
W Wegscheider, L Pfeiffer, M Dignam, A Pinczuk, K West and R Hull

1939 Recombination kinetics and intersubband relaxation in semiconductor quantum wires
M Grundmann, J Christen, M Joschko, O Stier, D Bimberg and E Kapon

1946 Spectroscopy of field-effect-induced superlattices
W Hansen, A Schmeller, H Drexler, J P Kotthaus, M Holland, G Tränkle, G Böhm and G Weimann

1953 Optical study of II-VI semiconductor nanostructures
Le Si Dang, C Gourgon, N Magnea, H Mariette and C Vieu

SECTION 5. BLOCH OSCILLATIONS AND ULTRAFAST PHENOMENA

1959 Terahertz Bloch oscillations in semiconductor superlattices
T Dekorsy, P Leisinger, C Waschke, K Köhler, K Leo, H G Roskos and H Kurz

1965 Coherent dynamics of exciton wavepackets in semiconductor heterostructures
J Feldmann, M Koch, E O Göbel, F Jahnke, T Meier, W Schäfer, P Thomas, S W Koch, H Nickel, S Luttgren and W Stolz

1972 Picosecond and femtosecond spectroscopy of highly excited charge carriers in semiconductors: theory
P Kocevar

1978 Theory of ultrafast phenomena in laser-excited semiconductors
D K Ferry

SECTION 6. COUPLED QUANTUM WELLS AND SUPERLATTICES

1983 Long-lived excitonic ground states in GaAs/AlAs coupled quantum well structures
A Zrenner, L V Butov and M Hagn

1989 Bloch and localized electrons in semiconductor superlattices
M Helm, W Hilber, T Fromherz, F M Peeters, K Alavi and R N Pathak

1994 Wannier-Stark effect in superlattices
C Hamaguchi, M Yamaguchi, M Morifuji, H Kubo, K Taniguchi, C Gmachl and E Gornik

1999 Optical spectroscopy on quantum wells and tunnel structures under conservation of angular momentum
G Hendorfer

SECTION 7. Si/SiGe HETEROSTRUCTURES

2005 UHVCVD growth of Si/SiGe heterostructures and their applications
B S Meyerson, K E Ismail, D L Harame, F K LeGoues and J M C Stork

2011 Photo- and electroluminescence in short-period Si/Ge superlattice structures
J Olajos, J Engvall, H G Grimmeiss, U Menczigar, M Gail, G Abstreiter, H Kibbel, E Kasper and H Presting

2017 Investigation of luminescence in strained SiGe/Si modulated quantum well and wire structures
Y Shiraki and S Fukatsu

SECTION 8. MICROCAVITIES

2025 Optical processes in microcavities
R E Slusher

SEMICONDUCTOR SCIENCE & TECHNOLOGY

Volume 9 Number 11S November 1994

8th Winterschool on New Developments in Solid State Physics Mauterndorf, Austria, 14–18 February 1994

SECTION 1. FRACTIONAL QUANTUM HALL EFFECT AND COMPOSITE FERMIONS

1853 The fractional quantum Hall effect in a new light
H L Stormer, R R Du, W Kang, D C Tsui, L N Pfeiffer, K W Baldwin and K W West

1859 Theory of the half-filled Landau level
N Read

1865 Inelastic light scattering in the regimes of the integer and fractional quantum Hall effects
A Pinczuk, B S Dennis, L N Pfeiffer and K W West

SECTION 2. MESOSCOPIC TRANSPORT AND CHAOS

1871 Interactions and transport in nanostructures
B Kramer, T Brandes, W Häusler, K Jauregui, W Pfaff and D Weinmann

1879 Dephasing by coupling with the environment, application to Coulomb electron–electron interactions in metals
Y Imry and A Stern

1890 Transport spectroscopy on a single quantum dot
J Weis, R J Haug, K von Klitzing and K Ploog

1897 Phase-breaking rates from conductance fluctuations in a quantum dot
C M Marcus, R M Clarke, I H Chan, C I Duruöz and J S Harris

1902 Chaotic transport and fractal spectra in lateral superlattices
R Fleischmann, T Geisel, R Ketzmerick and G Petschel

1906 Quantum transport and quantum chaos in antidot superlattices in a magnetic field
H Silberbauer, P Rotter, M Sührke and U Rössler

SECTION 3. LOW-DIMENSIONAL TUNNELLING

1912 Resonant tunnelling quantum dots and wires: some recent problems and progress
P H Beton, H Buhmann, L Eaves, T J Foster, A K Geim, N La Scala Jr, P C Main, L Mansouri, N Mori, J W Sakai and J Wang

1919 Low-dimensional resonant tunnelling and Coulomb blockade: a comparison of fabricated versus impurity confinement
M R Deshpande, E S Hornbeck, P Kozodoy, N H Dekker, J W Sleight, M A Read, C L Fernando and W R Frenksley

1925 Tunnelling spectroscopy of 0D states
J Smoliner, W Demmerle, E Gornik, G Böhm and G Weimann

Continued on inside back cover